Protein Crystallization Induced by Multivalent Metal Salts

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Andrea Sauter

aus Rottweil

Tübingen

2015

Tag der mündlichen Qualifikation: 20.7.2015

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Frank Schreiber

2. Berichterstatter: Prof. Dr. Thilo Stehle

Contents

Deutsche Zusammenfassung

0.1 Motivation & Einleitung

Proteine und ihre Wechselwirkungen in wässriger Lösung spielen eine zentrale Rolle für die biologischen Funktionen in jedem Lebewesen. Diese Arbeit beschäftigt sich mit der Beeinflussung dieser Wechselwirkungen durch Ändern von Kontrollparametern wie der Salzkonzentration oder der Temperatur, um das Phasenverhalten gezielt zu beeinflussen und Proteine zur Kristallisation zu bringen.

Die unerwünschte Aggregation oder Kristallisation von Proteinen spielt eine Schlüsselrolle für viele Krankheiten. Beispiele sind Sichelzellanämie, die durch Fasern aus sichelförmigen Hämoglobinmolekülen ausgelöst wird oder auch durch die Aggregation des Proteins γ-Kristallin im Glaskörper des Auges verursachter, altersbedingter grauer Star[70]. Einige neurodegenerative Erkrankungen wie etwa Alzheimer, Parkinson, die Creutzfeldt-Jakob-Krankheit, amyotrophe Lateralsklerose und Chorea Huntington rühren von einer irreversiblen Bildung von Proteinaggregaten, üblicherweise Fasern aus fehlerhaft gefaltetem Eiweiß, her[3,121,155]. Nachweislich können Salze die Faserbildung beeinflussen, wobei sowohl die Ionenstärke als auch die Art des Anions eine Rolle spielen[84].

Seit der ersten Hälfte des 20. Jahrhunderts wird Proteinkristallisation zur Aufreinigung von Proteinen eingesetzt. Beispiele sind eine Vielzahl an Serum-Albuminen, Enzyme und viele andere[62]. Die große Bedeutung von Salzen, Lösungsmitteln, dem pH-Wert und der Temperatur wurden schnell bemerkt und die Hofmeisterreihe dazu eingesetzt, um eine Übersättigung der Lösung zu erreichen. Die Wichtigkeit von Metallionen für die Proteinkristallisation wurde entdeckt, insbesondere an der Kristallisation von Insulin, die nur durch Zugabe von Zn^{2+} erreicht werden konnte[2,62]. Zudem enthalten über 90% aller pharmazeutischen Produkte bioaktive Wirkstoffe in kristalliner Form, in erster Linie, um Stabilität zu gewährleisten[47]. Sich langsam im Körper auflösende Kristalle ermöglichen eine kontinuierliche Bereitstellung des Wirkstoffes und stellen damit eine weitere Einsatzmöglichkeit von Proteinkristallen im medizinischen Bereich dar[70,204].

Ihre hauptsächliche Anwendung finden sie jedoch in der Strukturbiologie: Die dreidimensionale Struktur von Proteinen ist eng verknüpft mit ihrer biochemischen Funktion[127]. Daher ist die genaue Kenntnis von Proteinstrukturen essenziell für das Verständnis von grundlegenden physiologischen Prozessen und die gezielte Entwicklung von Medikamenten[124,127]. Gerade das Humangenomprojekt hat durch die Sequenzierung des menschlichen Genoms enorme Möglichkeiten für das Verständnis von durch Fehlfunktionen der Zelle verursachte Krankheiten und

für die Entwicklung neuer medikamentöser Therapien eröffnet. Die entsprechende Proteinstruktur zu allen kodierenden DNS-Regionen zu bestimmen ist daher eine große Herausforderung[29].

Weitere Möglichkeiten, Informationen über die Proteinstruktur zu gewinnen, sind Kernspinresonanz (nuclear magnetic resonance, NMR), Massenspektrometrie und Mikrokalorimetrie. Die universellste und erfolgreichste Methode zur Strukturbestimmung bleibt jedoch Röntgenbeugung. Am ehesten stellt NMR eine Alternative dar, sie ist jedoch auf kleine Makromoleküle mit weniger als 70 kDa beschränkt[127]. Zu den Anforderungen der Röntgenbeugungsmethode gehören dagegen relativ große und hochqualitative Einkristalle. Aus neu exprimierten Proteinen solche hochqualitativen Kristalle zu züchten ist jedoch keine einfache Aufgabe und stellt nach wie vor den hauptsächlichen Flaschenhals der Proteinstrukturbestimmung dar[127]. Die vielfältigen (teils durch die inhomogene Ladungsverteilung und die unregelmäßige Form der Proteine meist anisotropen) Wechselwirkungen von Proteinen untereinander und mit ihrer Umgebung, wie Wasserstoffbrückenbindungen, elektrostatische Wechselwirkungen oder Van-der-Waals-Kräfte, spielen eine entscheidende Rolle für die Proteinkristallisation[43,127] und machen es schwierig, diesen Prozess mit einfachen Modellen zu beschreiben. Ein weiterer wichtiger Punkt ist, dass die Neigung zur Kristallisation bei den meisten Proteinen die evolutionäre Fitness des Organismus beeinträchtigt. Zwar existieren Ausnahmen, in denen die Kristallisation von Proteinen vorteilhaft ist, in der Regel schadet dies jedoch dem Überleben der Zelle[41,42]. Daher haben sich die meisten Proteine im Laufe der Evolution dahingehend entwickelt, zumindest unter physiologischen Bedingungen schwer kristallisierbar zu sein[41].

In den letzten Jahren wurde in unserer Arbeitsgruppe entdeckt, dass die Wechselwirkungen zwischen Proteinen in Lösung durch “reentrant condensation” (RC) bei der Zugabe von Salzen angepasst werden können und weiteres vielfältiges Phasenverhalten wie flüssig–flüssig Phasenauftrennung (liquid–liquid phase separation, LLPS) oder Kristallisation ausgelöst werden kann[86,151,157,220,222,223,226,227]: Bei steigender Salzkonzentration tritt eine erste Grenze c^* auf, an der ein Übergang von einer klaren Lösung bei kleinen Salzkonzentrationen (Regime I) zu einer trüben Probe mit durch attraktivere Wechselwirkungen gebildeten Aggregaten auf (Regime II). Bei weiter steigender Salzkonzentration tritt eine zweite Grenze c^{**} auf, ab der die Proben wieder vollständig klar sind (Regime III). Diese Methode wird in dieser Arbeit weiterverfolgt, um die Bedingungen für Kristallisation zu optimieren. Ein wesentlicher Bestandteil der Arbeit besteht darin, die Kristallkeimbildung genauer zu beschreiben und unterschiedliche Mechanismen zu unterscheiden. Dies kann durch die schlechte Zugänglichkeit eines Keims oft nur indirekt geschehen. Die hier vorgestellte Arbeit liefert ein tieferes Verständnis dieser Prozesse in den untersuchten Modellsystemen. Die Möglichkeiten zur Anwendung in anderen Systemen werden diskutiert.

0.2 Ergebnisse

In diesem Abschnitt werden die Ergebnisse der Publikationen A-D zusammengefasst, die in wissenschaftlichen Zeitschriften erschienen oder erscheinen sollen (Publikation D). Sie entsprechen den Kapiteln 3-6. Abschnitt 0.2.5 fasst die Untersuchungen zur Übertragbarkeit auf andere Systeme aus Kapitel 7 zusammen.

0.2.1 Publikation A: Nicht-klassische Proteinkristallisation mit mehrwertigen Metallionen

In Publikation A wurde das Kristallisationsverhalten des Proteins Beta-Lactoglobulin (BLG) in Lösung mit Yttriumchlorid (YCl_3) oder Zinkchlorid ($ZnCl_2$) untersucht. Durch $ZnCl_2$ (oder auch durch das ebenfalls zweiwertige Salz Cadmiumchlorid ($CdCl_2$) kann eine erste Grenze c^* ausgelöst werden. Bei hohen Salzkonzentrationen kommt es jedoch nur zu einem teilweisen Wiederaufklaren ab einer als *Pseudo-*c^{**} bezeichneten Konzentration. Mittels Kleinwinkelröntgenstreuung (SAXS) wurde gezeigt, dass die effektiven Wechselwirkungen sich bei steigender Salzkonzentration von repulsiv im ersten Regime zu attraktiv verändern. Attraktive Wechselwirkungen dominieren auch im Bereich von *Pseudo-*c^{**}.

Die besten Bedingungen für Proteinkristalle hoher Qualität wurden nahe c^* und c^{**} (im Fall von YCl_3) bzw. *Pseudo-*c^{**} (im Fall von $ZnCl_2$) gefunden. Anders als für HSA beobachtet[223], findet Kristallisation zwar im gesamten zweiten Regime statt, die Kristallqualität nimmt weiter von den Grenzen entfernt jedoch deutlich ab. Unsere Strukturuntersuchungen zeigen auch, dass die Metallionen ein integraler Bestandteil des Kristallgitters sind. Sowohl Zn^{2+} als auch Y^{3+} können neue Proteinkontakte im Gitter durch Ionenbrücken erzeugen, wobei Yttrium im Vergleich zu Zink eine größere Zahl solcher Kontakte bildet.

Während nahe der unteren Grenze c^* die Kristallisation vermutlich klassisch abläuft, d.h. direkt aus der Lösung und ohne weitere Zwischenphasen, tritt nahe (*Pseudo-*)c^{**} vor der Kristallisation eine weitere Phase auf. Die Morphologie dieser Phase hängt stark von der Kristallisationstemperatur ab: Oberhalb einer Übergangstemperatur T_{tr} sind die betreffenden Proben klar und kristallisieren nicht[227]. Nahe T_{tr} besteht die Zwischenphase aus Proteinclustern. Deutlich unterhalb T_{tr} werden diese zu makroskopischen Aggregaten, die sich nach einiger Zeit zu einer dichten flüssigen Phase umwandeln. Kristallisation wird schließlich an der Grenzfläche zwischen dieser dichten und der umgebenden dünnen flüssigen Phase beobachtet. Unsere Beobachtungen mittels optischer Mikroskopie können jedoch nicht beantworten, ob es sich bei der Kristallkeimbildung um einen zweischrittigen Prozess handelt, bei dem die dichte Phase einen direkten Vorläufer des Kristalls darstellt oder ob nicht vielmehr heterogene Keimbildung der dünnen Phase an der Tropfenoberfläche stattfindet. Mögliche Szenarien werden anhand von Phasendiagrammen kolloidaler Systeme diskutiert.

0.2.2 Publikation B: Echtzeit-Beobachtung der Kinetik von nicht-klassischer Proteinkristallisation

(Dieser Abschnitt enthält übersetzte Auszüge eines für GIT Labor Fachzeitschrift geschriebenen Artikels)

In Publikation B wurde BLG in Lösung mit $CdCl_2$ untersucht. Wie im System mit $ZnCl_2$ wurde Kristallisation im gesamten zweiten Regime und etwas über die Grenzen hinaus beobachtet. Für die Strukturbestimmung gut geeignete Qualität hatten jedoch nur Kristalle, die aus Proben nahe der Grenzen c^* oder *Pseudo-*c^{**} gezüchtet wurden.

Direkt unterhalb c^*, wo die gesamte Probe klar ist oder etwas darüber, wo nach Zentrifugation ein klarer Überstand zurückbleibt, kann Kristallisation direkt aus der Lösung beobachtet werden. Im Bereich von *Pseudo-*c^{**} sind Proben trüb und unter dem Lichtmikroskop sind Proteinaggregate zu sehen (im Gegensatz zu Proben nahe c^*, bei denen auch in trübem Zustand die Aggregate zu klein für die Beobachtung sind). Unsere Beobachtungen und die Analyse des Kristallwachstums deuten darauf hin, dass die Aggregate eine entscheidende Rolle für die Keimbildung spielen.

Aufgrund der geringen Auflösung und Statistik der Beobachtung mit Lichtmikroskopie wird SAXS eingesetzt, um die strukturelle Entwicklung der Proben zu charakterisieren. Nahe c^* ähneln die Messkurven dem Formfaktor von BLG. Nach einiger Zeit tauchen Bragg-Peaks auf und zeigen die Kristallisation der Probe an. Es wurden keine Anzeichen für andere Strukturen als Lösung und Kristall gefunden. Im Bereich von *Pseudo-*c^{**} weichen die Messkurven stark vom Formfaktor ab. Es tritt zudem ein breiter Peak bei mittleren q-Werten auf, der mit zunehmender Zeit in seiner Intensität anwächst. Dies kann durch die Bildung von Proteinaggregaten mit einer bestimmten inneren Struktur erklärt werden. Zusätzliche Bragg-Peaks treten auf, deren Intensität zunimmt, während der breite Peak schließlich kleiner wird und verschwindet.

Um die Beziehung zwischen Aggregaten und Kristallen zu quantifizieren, wird die Fläche unter dem breiten Peak und die unter den Bragg-Peaks ausgewertet und als Funktion der Zeit aufgetragen. Es können zwei interessante Eigenschaften der Kristallisationskinetik beobachtet werden: Erstens fällt das Maximum in der Fläche unter dem breiten Peak mit einem lokalen Maximum der Kristallisationsrate (erste Zeitableitung der Fläche unter den Bragg-Peaks) zusammen. Dies ist charakteristisch für einen zweischrittigen Prozess, bei dem die Keimbildung aus einer Zwischenphase stattfindet und daher proportional zur Menge des Proteins in der Zwischenphase ist. Zweitens tritt, wenn der Großteil der Zwischenphase konsumiert wurde, eine starke Zunahme der Bragg-Peak-Fläche auf. Wir interpretieren dies als Wachstum der Kristalle in Kontakt mit der Lösung, in der zwar die Keimbildung weniger wahrscheinlich, aber das Wachstum der Kristalle durch die schnellere Diffusion, verglichen mit der in den Aggregaten, begünstigt sein könnte.

Wir verwenden Ratengleichungen basierend auf Prozessen mit klarer physikalischer Bedeutung: Zunächst bildet sich die Zwischenphase (Aggregate mit noch nicht kristalliner Struktur), bis diese eine stabile Konzentration erreicht. Dann bilden sich Kristalle in der Zwischenphase und wachsen dort langsam. Sobald das Volumen der Zwischenphase pro Kristall durch Keimbildung und Wachstum der Kristalle unter einen bestimmten Wert fällt, kommen die Kristalle in Kontakt mit der Lösung. Im letzten Schritt wachsen sie so, bis die freien Proteinmoleküle aufgebraucht sind. Ein Modell basierend auf diesen Annahmen kann die experimentell beobachteten kinetischen Merkmale mit guter Übereinstimmung reproduzieren.

Unsere Ergebnisse weisen daher auf eine entscheidende Rolle der Aggregate bei der Keimbildung sowie einen zweischrittigen Prozess hin. Außerdem demonstrieren sie eine nicht-invasive Methode für Echtzeitbeobachtungen der Kinetik von nicht-klassischen Wachstumsprozessen.

0.2.3 Publikation C: Zur Frage der zweischrittigen Keimbildung in der Proteinkristallisation

Aufbauend auf Publikation B soll in Publikation C die metastabile Zwischenphase im System BLG mit $CdCl_2$ und ihr Einfluss auf die Keimbildung und das Wachstum von Kristallen im oben beschriebenen zweischrittigen Prozess weiter untersucht werden. Während die Kleinwinkelstreukurven an anderen Punkten des Phasendiagramms im Bereich von *Pseudo-c*** sowie die zugehörigen Analysen den Befunden aus Publikation B stark ähneln, zeigt sich eine deutliche Abhängigkeit von der Salzkonzentration.

Serien von Proben nahe der oberen Grenze mit nur schwach variierter $CdCl_2$ Konzentration (33 mg/mL BLG mit 17, 18, 19 und 20 mM) zeigen klar, dass mit abnehmender Salzkonzentration größere aggregierte Bereiche gefunden werden und mehr Kristalle wachsen. Während die Keimbildungsrate anwächst, nimmt die Dauer der Wachstumsphase ab und die Kristalle werden, aufgrund der kleineren Menge an Protein pro Kristall, nicht so groß wie bei höheren Salzkonzentrationen. Unterhalb von 17 mM befinden sich Probenbedingungen tief im zweiten Regime, wo sich, aufgrund der starken Attraktion, massive Aggregate bilden und aus der Lösung ausfallen. Oberhalb von 20 mM $CdCl_2$ sind die Lösungen klarer und nur kleine Aggregate werden beobachtet. Diese Proben kristallisieren nicht oder nur sehr selten. Die Treibkraft für Kristallisation nimmt also beim Überqueren von *Pseudo-c*** ab.

Wenn wir SAXS-Kurven aus unterschiedlichen Regimen vergleichen wird deutlich, dass die Streuung im ersten Regime vom Formfaktor von BLG (Dimer) dominiert wird. Im zweiten Regime kommt ein Anstieg bei kleinen q-Werten, also großen Distanzen hinzu, der die Bildung von Aggregaten anzeigt. Im dritten Regime und im Bereich von *Pseudo-c*** tritt der schon in Publikation B als Struktursignatur des Kristallisationsvorläufers beschriebene breite Peak auf. Die aus Publikation B bekannte Analyse der SAXS-Daten wird in Publikation C erfolg-

reich unter anderen Kristallisationsbedingungen durchgeführt. Wieder zeigen sich der für einen Zweischrittprozess charakteristische Zusammenhang zwischen dem Maximum in der Fläche unter dem breiten Peak und dem lokalen Maximum in der Kristallisationsrate, sowie das Plateau in der Fläche unter den Bragg-Peaks als Funktion der Zeit. Da, wie bereits mit Lichtmikroskopie beobachtet, bei hohen Salzkonzentrationen nur wenige Kristalle wachsen, sind für die Zukunft Echtzeit-SANS-Messungen (größere Strahlbreite) geplant.

0.2.4 Publikation D: Zur Frage der Entwicklung metastabiler Proteinaggregate

Thema dieser Publikation ist die strukturelle Entwicklung von BLG-Aggregaten in Gegenwart von YCl_3. Es wurden Proben ausgewählt, in denen zuvor die in Publikation A beschriebenen nichtklassischen Kristallisationsprozesse beobachtet wurden. Um die Proben detaillierter beschreiben zu können, wurden diese bei Temperaturen oberhalb der Übergangstemperatur T_{tr} präpariert. Mittels einer Kombination aus SAXS, SANS und VSANS (Neutronenstreuung bei noch kleineren Winkeln, "very small angle neutron scattering") konnte die Struktur auf unterschiedlichen Längenskalen beschrieben werden, angefangen mit der Monomer–Monomer-Korrelation innerhalb größerer Partikel, über das BLG-Dimer und die Korrelation zwischen Proteinclustern bis hin zur räumlichen Anordnung der Cluster.

Werden die Proben unter T_{tr} gekühlt, bleibt die hierarchische Struktur im Wesentlichen erhalten. Es kommt jedoch zu einer Veränderung der inneren Struktur der Aggregate. So bildet sich die Monomer–Monomer-Korrelation vollständig zurück und die Cluster–Cluster-Korrelation wird stärker. Zudem weisen die Streudaten klar Punkte auf, an denen sich die Intensität der gestreuten Strahlung mit der Zeit nicht ändert. Diese isosbestischen Punkte weisen auf einen Prozess mit zwei Zuständen hin, die ein Gleichgewicht anstreben. Auf dieser Annahme beruhende Fits reproduzieren die Streudaten sehr gut. Wir schließen daraus, dass sich eine Zwischenphase ausbildet und dabei die zuvor vorhandenen, kleineren Proteincluster aufbraucht.

Diese Erkenntnisse könnten sich in der Zukunft als hilfreich für das Verständnis nichtklassischer Proteinkristallisation erweisen.

0.2.5 Übertragbarkeit der Kristallisationsmethode auf andere Systeme

Die Kristallisation des Proteins Beta-Lactoglobulin in den vorgestellten Publikationen beruht auf Nutzung des Phänomens der "reentrant condensation". In den darin beschriebenen Experimenten sowie in vorangegangenen Studien [221,227] wurde beobachtet und sich zu Nutze gemacht dass, durch die dort herrschende ideale schwache Attraktion, nahe der Grenzen c^* und c^{**} (oder *Pseudo-*c^{**} im Fall

der zweiwertigen Salze $CdCl_2$ und $ZnCl_2$) und im Fall von HSA auch nach einer flüssig–flüssig Phasenauftrennung[223] ideale Bedingungen für Proteinkristallisation zu finden sind.

Nun wurden eine Reihe anderer Protein–Salzsysteme auf den Einfluss von Kation und Anion auf RC, Ladungsumkehr und RC-induzierte Kristallisation getestet. Eine untere Grenze c^* wurde in allen untersuchten Systemen gefunden, eine obere nur in einem Teil. Die Variation des Kations hat, trotz deutlicher Unterschiede in der Ladungsdichte, nur geringe Auswirkungen. Bei Variation des Anions konnte festgestellt werden, dass zwei der getesteten Proteine nur mit Chloridsalzen eine obere Grenze aufweisen. Der Einfluss des Salzes ist jedoch noch nicht komplett verstanden.

Eine der Hauptanwendungen von RC ist die Kristallisation von Proteinen durch Anpassung der effektiven Wechselwirkungen zwischen den Proteinen. Tatsächlich konnten einige der getesteten Systeme erfolgreich kristallisiert werden, wenn auch nicht alle. Aufbauend auf den Erfahrungen mit BLG wurden — erfolgreich vor allem in Systemen mit HSA oder BLG — Proben für die Kristallisation gezielt nahe der Konzentrationsgrenzen präpariert.

Ein interessanter Ansatz für die weitere Entwicklung des Projektes ist die gezielte gentechnische Modifikation von Proteinoberflächen hin zu mehr negativ geladenen Aminosäuren, um Kristalle höherer Qualität mit der Metallsalz-Methode zu gewinnen.

0.3 Schlussfolgerungen und Ausblick

Im Folgenden werden die zentralen Ergebnisse dieser Arbeit zusammengefasst. Stärke und Vorzeichen der Wechselwirkungen zwischen negativ geladenen, globulären Proteinen können durch RC angepasst werden. Die schwache Attraktion an den Grenzen c^* und $(Pseudo\text{-})c^{**}$ bieten ideale Bedingungen für die Kristallisation der Proteine, wie frühere Experimente zeigten und wie es auch in den in dieser Arbeit betrachteten Systemen bestätigt werden konnte.

Die Röntgenstrukturanalyse an in Gegenwart von $ZnCl_2$ und YCl_3 gezüchteten BLG-Kristallen haben eine hohe Auflösung und wurden in der Proteinstrukturen-Datenbank PDB hinterlegt. Die Metallionen sind ein integraler Bestandteil des Kristallgitters und beide Kationen bilden durch Ionenbrücken neue Kontakte zwischen Proteinen.

Während bei der Kristallisation nahe der ersten Grenze c^* bislang keine Anzeichen für einen nichtklassischen Prozess beobachtet wurden, treten nahe der oberen Grenze $(Pseudo\text{-})c^{**}$ zusätzliche Phasen vor der Kristallisation auf. Abhängig von der Temperatur wurden eine dichte flüssige Phase oder Aggregate beobachtet. Möglich sind sowohl ein zweischrittiger Keimbildungsprozess, als auch die parallele Bildung von Kristallen und Aggregaten bzw. dichter flüssiger Phase. Da uns diese Unterscheidung nicht für alle Systeme möglich war, bezeichnen wir alle Kristallisationsprozesse, bei denen weitere Phasen außer der ursprünglichen

Lösung und dem Kristall als Endzustand auftreten, als "nonclassical" und tatsächliche Zwei- oder Mehrschrittprozesse in der Kristallkeimbildung als "two- / multi-step nucleation".

BLG in der Gegenwart von YCl_3 wurde weiter unter Variation der Temperatur untersucht. SAXS, SANS und VSANS Messungen an Proben ober- und unterhalb der Übergangstemperatur zeigen Strukturen auf unterschiedlichen Längenskalen und ihre temperaturabhängigen Veränderungen. Eine Zweiphasen-Analyse reproduziert die Daten sehr gut.

Im System BLG+$CdCl_2$ gelang es über Beobachtung der Kristallisationskinetik mittels Echtzeit-SAXS, klare Hinweise auf einen Mehrschrittprozess zu finden. Aufgrund der experimentellen Beobachtungen schlagen wir den folgenden Ablauf vor: In den Aggregaten, die direkt nach der Probenpräparation entstehen, formt sich eine noch nicht kristalline Vorläuferstruktur, aus der sich anschließend Kristalle bilden. Diese wachsen zunächst eher langsam. Erst, wenn genug Aggregatphase durch Keimbildung und Wachstum der Kristalle aufgebraucht ist, kommen die Kristalle in Kontakt mit der umgebenden, weniger dichten Lösung. Während die Keimbildung in dieser deutlich unwahrscheinlicher ist als in der dichten Aggregatphase, ist das Wachstum der Kristalle hier aufgrund der ungehinderten Diffusion schneller. Ein auf diesen Annahmen aufbauendes Ratengleichungsmodell reproduziert die beobachtete Kinetik gut.

Weitere Experimente am System BLG+$CdCl_2$ zeigen, dass mit steigender Salzkonzentration im Bereich von *Pseudo-c*** sowohl die Menge an Aggregat, als auch die Anzahl der beobachteten Kristalle abnimmt (die dann, aufgrund der größeren verfügbaren Menge an Protein pro Kristall, größer werden können).

Für die Zukunft sind SANS-Messungen in Echtzeit geplant, da hierbei kein Strahlschaden auftritt und durch die höhere Strahlgröße an mehr Kristallen gemessen werden kann. Ebenso ist geplant, vergleichbare Messungen und Analysen für andere Protein-Salz-Systeme durchzuführen, etwa an HSA in Gegenwart verschiedener Salze.

Die Methode, Proteine mittels Metallsalzen zur Kristallisation zu bringen, ist ein vielversprechender Ansatz für weitere Systeme. Durch gezielte Modifikation der Oberflächengruppen von Proteinen könnten besser streuende Kristalle gezüchtet werden.

Abstract

Proteins and their interactions with each other are of interest for fundamental research as well as numerous applications in biology, biochemistry or medical and pharmaceutic industry. In recent years, research in our group has shown that trivalent metal salts can cause a rich phase behavior in aqueous protein solutions, including reentrant condensation, metastable liquid–liquid phase separation and crystallization. The topic of this thesis is on the one hand the influence of control parameters in these systems on protein aggregation and especially crystallization. On the other hand, we focus on the discrimination between classical and non-classical pathways of crystal nucleation and growth. We study mainly the model protein β-lactoglobulin (BLG) in the presence of the divalent salts zinc chloride and cadmium chloride and the trivalent yttrium chloride.

It has been observed that trivalent salts can cause a reentrant condensation in protein solutions: below a certain salt concentration c^* and above a second one, c^{**}, samples are clear, indicating stable solutions. At intermediate salt concentrations, samples become turbid due to aggregation. The divalent salts $CdCl_2$ and $ZnCl_2$ cause a first boundary in BLG solutions and a partial re-clearing at a salt concentration we denote as *pseudo-*c^{**}. Using SAXS, we show that the effective interactions become less repulsive approaching c^* and finally attractive. Although the attractive interaction becomes weaker when *pseudo-*c^{**} is crossed, repulsion is not reached again. The weak attraction close to the boundaries provides optimal conditions for crystallization. Depending on salt concentration and temperature, different pathways are found. We discuss possible scenarios based on the phase behavior of colloidal systems with short-range attractions. Structure determination by X-ray diffraction shows that the cations are an integral part of the lattice and mediate new contacts by the formation of ion bridges.

While it was not possible to distinguish if the observed intermediate phases are a direct precursor for crystals or merely act as an agent for heterogeneous, classical nucleation in the system with $ZnCl_2$, real-time SAXS measurements of BLG in the presence of $CdCl_2$ provide evidence for a multistep process around *pseudo-*c^{**}: in the SAXS data, a broad peak forms first, indicating an intermediate structure, followed by Bragg peaks. To quantify the relationship between the crystals and their possible precursor, the area under the Bragg peaks and under the additional broad peak were evaluated and plotted as a function of time. It can be observed that the maximum amount of the intermediate coincides with the steepest increase of the crystalline state. We propose a first step in which a non-crystalline, ordered structure is formed within metastable aggregates, followed by crystal nucleation in this precursor phase. Our observations further suggest that

the number of crystals increases strongly in the beginning, but their growth speed is low inside the aggregates that have a high viscosity which hinders the diffusion of protein molecules. After enough aggregate material has been consumed by crystal nucleation and growth, the crystals are in contact with the dilute solution and grow faster. The number of crystals, however, stays almost constant in this stage. Based on these assumptions, a rate equation model was applied which reproduces the experimentally observed kinetics well. Increasing the salt concentration in the vicinity of *pseudo-c*** leads to a reduction of aggregates and to fewer crystals that grow larger due the larger amount of available proteins per crystal.

The nonclassical crystallization pathway described above suggests that the structure of the intermediate phase may be crucial for the subsequent crystal nucleation. We thus performed a study on the structure of potential precursor phases in the system of BLG in the presence of YCl_3 on different length scales by a combination of SAXS, SANS and VSANS. The monomer–monomer correlation within larger compounds, the scattering of dimers as well as the correlation between clusters and their spacial arrangement can be identified. When the system is cooled below a certain transition temperature, changes in the internal structure of the intermediate phase can be observed. Moreover, time-dependent SAXS measurements show clear isosbestic points, allowing to describe the kinetics of the structural evolution of the intermediate phase by a two-state model.

Experiments on other protein–salt systems suggest that our work can be, to some extent, generalized. For some of the systems, structure determination by X-ray diffraction has been performed successfully. In all cases, crystal contacts are mediated by ion bridging. Tuning the protein–protein interactions by the addition of multivalent metal salts provides a method for the growth of potentially high-quality crystals.

Part I

Fundamentals

Chapter 1

Introduction

Interactions between proteins in aqueous solution are a fundamental process in all living beings. They are influenced by a variety of control parameters such as salt concentration, temperature and pH of the environment. The understanding and control of protein interactions and the resulting phase behavior has a large impact on applied science.

The topic of this thesis is tuning interactions in protein solutions by varying the salt concentration, salt type and temperature towards the desired phase behavior and especially crystallization. Crystallization itself is further investigated in order to determine the pathways it follows.

This introduction starts with a motivation of this work and an insight into the basic principles of phase transitions (Preliminary remarks, Section 1.1). Subsequently, the basics of crystallization (Section 1.2) and ion-induced interactions in protein systems (Section 1.3) will be introduced. In the end, an outline of the subsequent parts of this thesis is given.

1.1 Preliminary Remarks

1.1.1 Motivation

Aggregation or crystallization of proteins is of interest due to many reasons. The undesired aggregation of proteins in the body plays an important role in many diseases. Examples include sickle cell anemia that is caused by sickle hemoglobin molecules forming polymer fibers within the red blood cells or age-related cataract caused by the aggregation of the protein γ-crystallin in the vitreous fluid of the eye[70]. Several neurodegenerative diseases such as Alzheimer's Dementia, Creutzfeldt-Jakob disease, Parkinson's disease, amyotrophic lateral sclerosis and Huntington's disease are due to the irreversible formation of protein aggregates, usually fibrils containing misfolded protein[3,121,155]. It has been shown that salts can modulate the kinetics of fibril formation and that both ionic strength and nature of the anion play a role[84].

Since the first half of the 20^{th} century, protein crystallization was applied for purification of proteins. Examples are various serum albumins, several enzymes and many more[62]. In time, the importance of salts, solvents, pH and temperature were noted and the Hofmeister effects were applied to tune interactions and

optimize conditions for protein crystallization. The importance of metal ions for protein crystallization was discovered (in particular for the crystallization of insulin where Zn^{2+} has proven to be crucial)[2,62]. Furthermore, over 90% of all pharmaceutical products contain bioactive drugs in a crystalline state, mainly for reasons of stability[47]. The continuing release of drugs, for example insulin, provided by the slow dissolution of crystals in the body is beneficial, too[70,204].

The main application of protein crystals, however, is found in structural biology. The three-dimensional structure of proteins is closely linked to their biochemical function[127]. Therefore, the precise visualization of protein structures is essential for the understanding of fundamental physiological processes but also for rational drug design[124,127]. Especially in the era of the human genome project, the sequencing of the human genome has offered enormous opportunities for the understanding of human diseases that are due to abnormal functioning of cells and the development of drug therapies[40]. To solve the protein structures corresponding to all coding DNA regions is therefore a desired yet challenging task[29].

Although information on protein structure can be gained by nuclear magnetic resonance (NMR), mass spectroscopy and microcalorimetry, the universal and most successful tool for structure determination still is X-ray diffraction. The second most used method, NMR, is restricted to small macromolecules below 70 kDa[127]. Requirements for X-ray diffraction studies include relatively large and highly perfect single crystals. However, growing such high quality crystals from freshly expressed proteins is no easy task, remaining the major bottleneck in protein structure determination[127]. Despite the important fields of application, protein crystallization is often performed by trial and error approaches[127]. Numerous interactions (partially anisotropic due to the inhomogeneous charge pattern and irregular shape) of proteins with each other and with their environment such as hydrogen bonds, electrostatic interactions, or van der Waals interactions play a decisive role in protein crystallization[43,127], making it difficult to describe this process by easy models. Attention should also be paid to the fact that for most proteins, a low crystallization propensity increases the evolutionary fitness of the organism. Although there are examples of protein crystallization featuring survival benefits, this is rather exceptional and crystallization *in vivo* usually compromises the viability of the cell[41,42]. Thus, in the process of evolution, most proteins became hard to crystallize at least at physiological conditions[41].

Recent work has shown that we can tune protein interactions in solution inducing a so-called reentrant-condensation and related phase behavior such as liquid–liquid phase separation (LLPS) or crystallization by the addition of multivalent salts[86,151,157,220,222,223,226,227]. This method is further pursued in order to optimize the conditions for crystallization and especially to characterize and distinguish different mechanisms of nucleation and crystal growth. These are still not completely understood and multiple pathways are currently discussed in the community. The work in this thesis provides a deeper understanding of these processes in the investigated model systems.

1.1.2 Remarks on Phase Transitions

In the following, a few thermodynamic principles of the topics of this thesis will be discussed. The Helmholtz free energy F is useful for the calculation of thermodynamic properties in systems with fixed temperature, number of particles, and volume. If the pressure instead of the volume is held constant, the enthalpy H is used (and minimized in an equilibrium state of a system with pressure reservoir). The most commonly used free energy for the description of systems of globular proteins, however, is the Gibbs free energy or free enthalpy G for processes at constant temperature T and pressure P, where

$$G = U - TS + PV = H - TS \tag{1.1}$$

with the volume V, the entropy S, and the internal energy U[70].

Thermodynamic stability is an important concept predicting how a system will react to perturbation. A system is referred to as "stable" if the energy is at its global minimum and the system will remain in its current state. A "metastable" system translates to a local minimum. To reach the most stable state, an energy barrier has to be overcome. If the system is in an unstable state, the slightest fluctuation of a thermodynamic variable is sufficient to change into another, more stable state[149].

A phase is characterized by certain properties of a material, for example its state of matter, ferromagnetism or superconductivity[139]. If a phase becomes unstable in the given thermodynamic conditions, a phase transition takes place[139]. In order to be in equilibrium, temperature, pressure, and chemical potential of these two phases have to be the same[70]. A distinction is drawn between first-order and second-order phase transitions. In first-order transitions, thermodynamic potentials such as G are continuous but their first derivatives, e.g.

$$S = -\left(\frac{\partial G}{\partial T}\right)_P \qquad V = \left(\frac{\partial G}{\partial P}\right)_T \qquad H = \left(\frac{\partial (G/T)}{\partial (1/T)}\right) \tag{1.2}$$

show discontinuities and latent heat is associated to them. In second-order transitions, the potentials and their first-order derivatives are continuous while some second-order derivatives approach infinity or are reduced to zero. This classification can be extended to higher-order phase transitions[139]. Two- and higher order phase transitions are not associated to latent heat and, in contrast to first-order transitions, are accompanied by a change in symmetry (the liquid/gas critical point is an exception).

The van der Waals equation of state provides a simple picture of the gas–liquid transition[37]. It is shown in Equation 1.3 with the number of particles N, the Boltzmann constant k_B and the constants a and b. a takes into account the

attractive interactions and b the excluded volume of the particles.

$$\left(P + \frac{N^2 a}{V^2}\right)(V - Nb) = Nk_B T \tag{1.3}$$

Introducing the molar volume $\tilde{V}$ and the molar quantities $\tilde{a}$, $\tilde{b}$

$$\tilde{V} = \frac{N_A V}{N} \qquad \tilde{a} = N_A^2 a \qquad \tilde{b} = N_A b, \tag{1.4}$$

the van der Waals equation can also be written as the following expression for the pressure:

$$P = \frac{N_A k_B T}{\tilde{V} - \tilde{b}} - \frac{\tilde{a}}{\tilde{V}^2}. \tag{1.5}$$

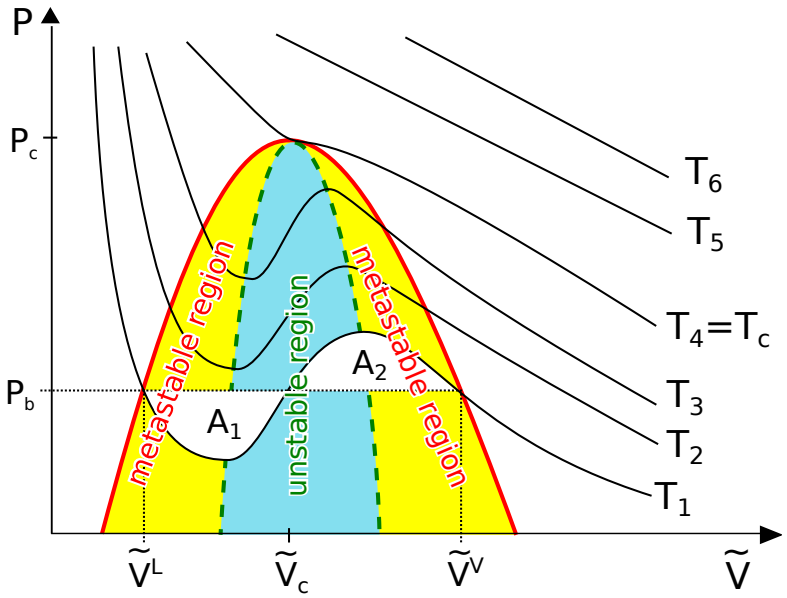

Figure 1.1: *Van der Waals isotherms at different temperatures (P as a function of $\tilde{V}$) and Maxwell construction. Image adapted from Ref. [37] and modified.*

Below a critical temperature (T_c), an isotherm $P(\tilde{V})$ of a van der Waals system has a local minimum and a local maximum as shown in the curve marked as T_1 in Figure 1.1. At molar volumes $\tilde{V} < \tilde{V}^L$, the system is in its liquid state, at $\tilde{V} > \tilde{V}^V$, it is in the vapor state. Due to the rule of stability, $(\delta P/\delta \tilde{V})_T$ has to be negative. Thus, the region marked as unstable in Figure 1.1 is not realized in physical systems and the system rather splits into two phases. In reality, the pressure remains constant between $\tilde{V}^L$ and $\tilde{V}^V$, denoted as P_b, according to the principle of equal areas (Maxwell construction)[33,37]: since the variation of the

chemical potential μ along an isotherm is given by

$$d\mu = \tilde{V} dP \tag{1.6}$$

and $P^\alpha = P^\beta$, $\mu^\alpha = \mu^\beta$ for two phases α, β at the equilibrium,

$$\int_{\tilde{V}^\alpha}^{\tilde{V}^\beta} \tilde{V} dP = \mu^\beta - \mu^\alpha = 0 \text{[37]}. \tag{1.7}$$

The construction shown in Figure 1.1 satisfies Eq. 1.7 as well as $P_b(T_1) = P^V(T_1) = P^L(T_1)$ with P^V the pressure of the vapor and P^L the pressure of the liquid phase[37].

The regions marked as metastable in Figure 1.1 represent a supersaturated vapor and an overexpanded liquid. They can be reached under special conditions, not violating the rule of stability, but convert to the two-state system if a sufficient disturbance occurs[98]. The outer line (red in Figure 1.1) is called coexistence line or binodal, the inner line (green, dashed) is called spinodal[139]. If the latter is crossed, the system becomes unstable and immediately phase-separates. In the metastable region between spinodal and binodal, an energy barrier has to be overcome in order to achieve phase separation[85].

The critical isotherm $T_4 = T_c$ has a point of inflection at $\tilde{V}_c$ and no extrema. At this critical point, the two phases become identical[10]. Above T_c, the isotherms are monotonically decreasing functions such as T_5 and T_6. Only one fluid phase exists at a given pressure for these temperatures[37].

These thermodynamic principles are also applicable to solutions, where the multi-component analog of the gas–liquid transition can be observed: a dilute–concentrated transition, also called LLPS or oiling out[10,102,142,171]. Particles in solution can — at small concentrations where they do not interact much — be understood in analogy to the gas. In both scenarios, a new phase in which the molecules are present at much higher concentration forms. This concentration is usually only slightly below that in a crystal[171]. In phase diagrams for protein solutions, usually the volume fraction Φ is plotted as one of the variables.

In a regular binary solution with two components A and B at concentrations c_A and c_B and G_A^0 and G_B^0 being the Gibbs free energies of the pure substances, the Gibbs free energy of the solution is

$$G = c_A G_A^0 + c_B G_B^0 + u c_A c_B + RT[c_A ln(c_A) + c_B ln(c_B)] \tag{1.8}$$

where R is the ideal gas constant[139]. The third term represents the energy of interaction of particles A with particles B (with internal energy per unit mass or volume u), the last term the entropy contribution of mixing. Assuming

$$x = c_A \qquad c_A + c_B = 1 \quad \Rightarrow \quad c_B = 1 - x, \tag{1.9}$$

the metastability limit $\delta^2 G = 0$ can be written as

$$-2ux(1-x) + RT = 0 \tag{1.10}$$

which gives the spinodal curve shown in Figure 1.2 as dashed line[139]. Above the critical point at $x = 0.5$ and $T_c = u/RT$, A and B are in a homogeneous solution at all mixing ratios. Inside the spinodal, they separate. Using that the chemical

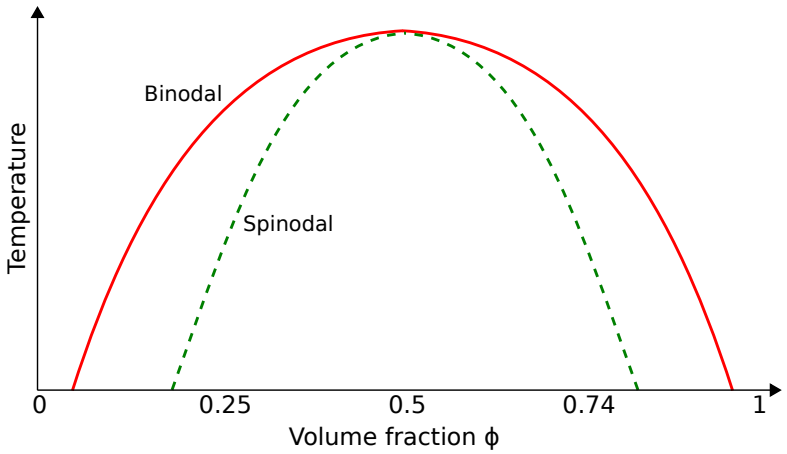

Figure 1.2: *Binodal and spinodal for a regular solution. The solution is metastable between both curves and becomes unstable inside the spinodal. Adapted from Ref. [139] and modified.*

potential of a constituent is the same in the mixed and the separated system, the coexistence curve (binodal) can be found[139]:

$$\frac{1-x}{x} = exp\left[\frac{(1-2x)u}{RT}\right]. \tag{1.11}$$

Between binodal and spinodal, the system is metastable. Put into a metastable two-phase state, e.g. due to a temperature change, a second phase nucleates within a formerly single-phase system, being energetically favorable[147]. Solutions in this region will phase separate via formation and growth of droplets after a sufficient fluctuation[139].

Depending on the range of interaction, very different types of $T(\Phi)$ phase diagrams can be obtained. Colloidal systems are useful models here as presented by Anderson and Lekkerkerker[7]. For a purely hard sphere system, a phase diagram as shown in Figure 1.3a is found. A fluid (F) and a crystalline (C) phase are found with a volume fraction range of coexistence (F+C). Introducing long-range attractions (e.g. model for argon), the phase diagram dramatically changes as shown in Figure 1.3b and a third phase, liquid (L) appears. If the attractions are short-ranged such as in a typical protein solution, the (L+L) transition becomes metastable, see Figure 1.3c. A similar phase diagram is found for short polymers[7].

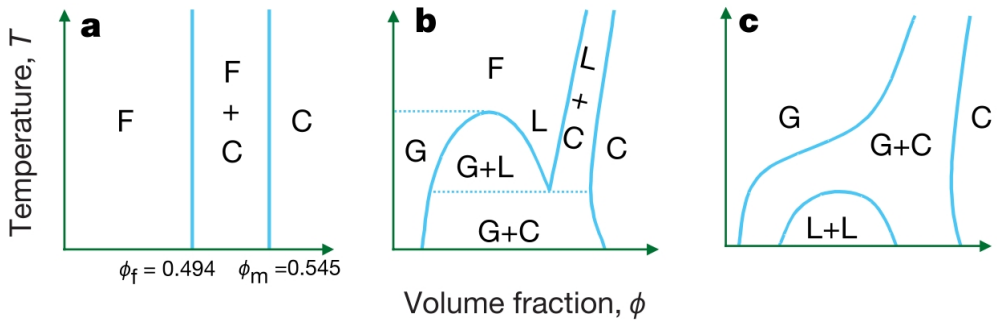

Figure 1.3: *Phase diagrams for a) hard spheres. b) hard spheres with long-range attractions. c) a "protein" system where the attraction is short-ranged. The liquid–liquid transition becomes metastable. Reproduced from Ref. [7].*

1.2 Nucleation and Crystal Growth

Before nonclassical crystallization processes can be discussed, this section introduces "classical" crystallization itself. In the end, the peculiarities of *protein* crystallization are discussed and open questions specified.

Crystallization starts from dissolved atoms, molecules or ions, its thermodynamic driving force being the supersaturation of the solution [34]. The relative supersaturation s is defined as

$$s = \frac{c}{k_{sp}} \tag{1.12}$$

with c the concentration of the species and k_{sp} the equilibrium molecular solubility product [34].

Since the magnitude of supersaturation, however, is not the only driving force in the control of crystallization, it can only be a first thermodynamic approach to the complex problem of crystallization [34]. Supersaturation can be translated into attraction of molecules or ions leading to a non-statistical distribution of these building blocks [34]. The relative supersaturation is connected to the change in chemical potential $\Delta\mu$ and therefore to the free enthalpy of the crystallization process:

$$\Delta\mu = -k_B T \cdot \ln(s) \tag{1.13}$$

Supersaturation can be achieved in different ways, for example by reactions generating the respective soluble, by a temperature or pressure jump or by changing the pH [34].

Once the solution is supersaturated (i.e. $s > 1$), crystals can grow, but need a so-called nucleus as starting point [34]: the minimum amount of the new phase that can exist independently [47]. Nucleation is the local generation of a distinct, more stable thermodynamic phase — in the case of crystal nucleation solid — from a metastable one [202]. A distinction is drawn between homogeneous nucleation, where the nucleus is formed from solution as soon as a critical supersaturation is given and heterogeneous nucleation, where surfaces or contaminants like dust

particles or crystal seeds act as starting point[34]. Heterogeneous nucleation is favored if the interface energy of the crystal being in contact with the substrate is lower than the interface energy of the crystal being in contact with the free solution, which is the case especially for impurities very similar to the crystal that can be used for crystal seeding[34]. Already a single foreign particle down to molecular size could act as an agent for heterogeneous nucleation and even in very pure solutions, heterogeneous nucleation occurs along surfaces such as the container wall[34,43]. This circumstance makes it difficult to prove that nucleation is ever completely homogeneous. Prior filtration often significantly reduces the number of crystals[43]. A special case of heterogeneous nucleation can be observed on surfaces with matching lattice where nucleation is especially fast. A manifest example is the formation of a new layer on a growing crystal[171]. Although heterogeneous nucleation is the normal case, most of the knowledge about microscopic dynamics of nucleation has been derived from simulation which usually models homogeneous nucleation of very simple systems such as noble gases[171]. In colloidal systems, the particles can also be monitored directly.

Crystallization often is perceived as counter-intuitive in terms of entropy. As stated before, T and P are usually held constant in protein crystallization experiments and thus the thermodynamic potential that is minimized is the Gibbs free energy. The equation for describing the Gibbs free energy change of a crystallization process is

$$\Delta G = \Delta H - T\Delta S_{protein} - T\Delta S_{Solvent}. \tag{1.14}$$

$\Delta S_{protein}$ describes the (usually negative) change in entropy caused by the attachment of protein molecules to the crystal. On the one hand, there is a loss of entropy due to constrained rotational and translational degrees of freedom[70]. On the other hand, there is a gain in vibrational entropy associated to new bonds at molecular contacts[70]. $\Delta S_{Solvent}$ describes the entropy change caused by the water shells formed around proteins due to protein–water interactions. When the crystal is formed, these water structures have to reorder. This term can be either positive or negative but ΔG has to be negative in order to enable crystallization[70]. Vekilov *et al.* showed that this trapping and release of water is a major factor in protein crystallization, using three example systems[205]: human hemoglobin C has a high positive enthalpy. Crystallization is only possible due to a large entropy gain originating from the release of water. Apoferritin has a crystallization enthalpy close to zero and the main driving force for crystallization is the entropy gain due to the release of water, too. For lysozyme the entropic effect is negative. Its crystallization is driven by the negative enthalpy. Even colloidal systems of hard spheres can crystallize at high volume fractions, purely driven by the increase in entropy due to the larger accessible volume for local movements in the lattice[55].

In the following, classical and nonclassical pathways of crystal nucleation will be discussed.

1.2.1 Classical Nucleation Theory

The classical nucleation theory (CNT) is the most common formalism to analyze nucleation phenomena[34]. Its main concepts on the generation of nuclei of a new phase were derived by Gibbs[60,61] (who considered the condensation of vapor to liquid) and later refined by Volmer and Weber[210] (1926) and Becker and Döring[18] (1935)[56,197,202]. Ignoring lattices and therefore also defects, it is essentially a theory for the nucleation of fluid phases[171].

In general, large aggregates of a new phase tend to grow further while small ones tend to dissociate. This can be explained by their surface to volume ratio: there are not enough intermolecular bonds in small aggregates to compensate for the decrease in entropy associated with growth. For a spherical cluster with radius r, this is included by a surface term proportional to r^2, taking into account the free energy cost of creating a vapor–liquid / solid–liquid interface[11]. In large aggregates, however, a volume term proportional to r^3 dominates. Thus, a free energy barrier has to be overcome that is formed by the usually positive surface term that hinders the formation of a new phase and the volume term driving the reaction that includes the (negative) chemical potential difference between initial and final state[34,43]. The barrier height is given by the change in Gibbs free energy ΔG (Equation 1.15, Figure 1.4), where ρ is the density of the bulk liquid and γ the (vapor–liquid / solid–liquid) interfacial free energy density[11].

$$\Delta G = \frac{4\pi r^3 \rho \Delta\mu}{3} + 4\pi r^2 \gamma \tag{1.15}$$

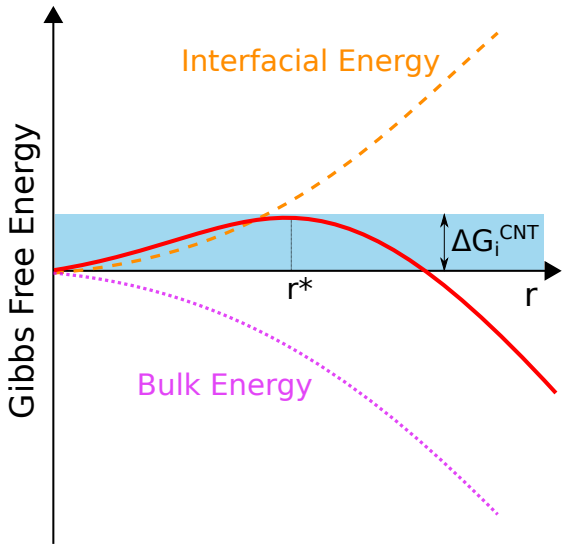

Figure 1.4: *The nucleation barrier in $\Delta G(r)$ is generated by a positive surface term $\propto r^2$ and a negative volume term $\propto r^3$. At the critical cluster size r^*, both terms balance each other. For $r > r^*$, further growth is favored. Image adapted from Ref. [34] and modified.*

The maximum of $\Delta G(r)$ corresponds to the so-called critical cluster size. From this radius r^* on, the energy gain for the phase transition is larger than the loss in surface energy for further growth of the so-called critical crystal nucleus[34,47]. Critical cluster sizes for water were found to consist of 20-35 molecules, depending on the temperature, and of 24-36 for pentanol[34]. Critical nucleus sizes for proteins were reported to range between five and 50 monomers[43]. However, these sizes of the critical crystal nucleus also depend on the shape and structure of the nucleus and in the CNT, a spherical solid cluster is always assumed[34].

The classical homogeneous nucleation rate J is given by[34]

$$J = K \cdot exp\left(\frac{-\Delta G}{k_B T}\right). \tag{1.16}$$

K is a kinetic prefactor. It has been modified over the years — after it was found to be inconsistent — but real nucleation rates still differ from the predicted ones by orders of magnitude and especially show a systematic dependence on temperature, indicating that the real crystallization scenario is more complicated. This can be corrected by empirical functional factors but a physical explanation is missing[34].

CNT has six assumptions (and for each of them situations can be found where they are not fulfilled). It assumes that nucleation is a one-step process (1). The nucleus is assumed to grow one monomer at a time (2) and the nucleation rate does not depend on the (for example thermal) history of the sample (3). Nucleation occurs at a saddle point, i.e. a minimum in (Gibbs) free energy and a maximum in other thermodynamic potentials (4) and the only source of slow kinetics is the nucleation barrier (5). Finally, the crystal lattice is neglected (6)[171].

Although CNT is still a useful model for the crystal nucleation of simple liquids or many ionic crystals, e.g. the crystallization of sodium chloride from the melt, it often comes to its limits in more complex systems[171]. This will be discussed in the following section.

1.2.2 Nonclassical Crystallization Processes in General

The term "nonclassical crystallization" covers numerous phenomena that deviate from the classical nucleation view in one way or another: biominerals that show single crystalline behavior in scattering experiments and have an unusual or complex morphology are hard to explain in the classical picture of a single crystal as a solid with a characteristic, rigid lattice with defined angles between certain faces[34]. The same is true for mesocrystals (colloidal crystals built up from smaller crystals) that usually scatter X-rays like a single crystal and can possess any outer shape including curvatures[34]. Furthermore, CNT cannot predict the dependence on temperature and the absolute values of critical supersaturations for various substances such as water, alcohols or alkanes with high molecular weight. The key source of these discrepancies is not entirely clear up to now[34].

Clusters often play an important role in crystallization and lead to nonclassical

behavior in the sense that a nucleus does not grow particle by particle but by coalescence of clusters of many particles[58]. A special case of nonclassical crystallization involves a two-step (or multistep) mechanism of crystal nucleation, i.e. an intermediate state, consisting of clusters, aggregates or a dense liquid phase, forms first from the initial liquid phase. In a second step, this intermediate further transforms into the final crystalline phase. Multiple intermediate states are possible, too. The idea that the initial and final state are not the only ones goes back to Ostwald's step rule which was formulated in the end of the 19^{th} century[88,136]. It states that a system passes from one state to the one with the next lowest free energy until it reaches the final state at the global minimum[106]. A novel two-step mechanism of crystal nucleation was initially proposed for protein systems and could be generalized for other molecules[62]. It has been proposed for various systems such as Ca^{2+} salts and similarly weakly solvated species where the formation of clusters and amorphous intermediates prior to crystallization can be observed[34]. The lower the solubility of a species, the higher is the probability of finding amorphous precursor stages because a lower solubility increases the apparent supersaturation, but also stabilizes amorphous nanoparticles against redissolution[34]. Depending on the free energy landscape, a system can follow a one-step route to the final crystal or a sequential process of structural and compositional changes. The intermediate phase closest to the initial phase in free energy (usually the least dense one) is formed first and transforms step-wise to further phases[34,171]. However, this is not always true since this transformation can be altered by surface properties and impurities[171].

The precursor can also be a liquid phase. LLPS can be observed in protein solutions and is also known to occur in systems of mineral salts such as $CaCO_3$[34]. Although the dense liquid phase accelerates crystallization in many systems, it can also hinder it under certain circumstances: if the liquid is arrested (very slow diffusion), then its formation can inhibit crystallization. If the "intermediate" phase is too stable, then it persists and no crystals form[171]. On the other hand, a not stable enough intermediate only forms transiently in microscopic amounts. It has to persist long enough for crystals to nucleate in it.

A general problem for the understanding of experimental results from molecular and ionic systems is that it is impossible to observe the nucleus directly[171]. Colloidal systems are, due to the large particle size and the slow dynamics, easier to access. First indications for a two-step process in colloidal hard spheres were found using time resolved laser light Bragg scattering[168]. Simulations by Kawasaki and Tanaka[88] showed that a supercooled liquid may not be spatially homogeneous as commonly expected, but has a medium-ranged structural order and this fluctuating order can promote crystal nucleation. This might also be a reason for the large discrepancy between estimations of the nucleation rate based on CNT and experimental observations[88]. In 2014, Peng *et al.* and Tan *et al.* directly followed the nucleation from the intermediate phase in a colloidal system by single-particle resolution microscopy[141,190]. Figure 1.5 shows structures of crystal precursors and nuclei in the early stage of crystallization. The precursors can be relatively

ordered structures with different symmetries. The work provides details on the structural aspects of two–step nucleation [67,190]. Although there are decisive dif-

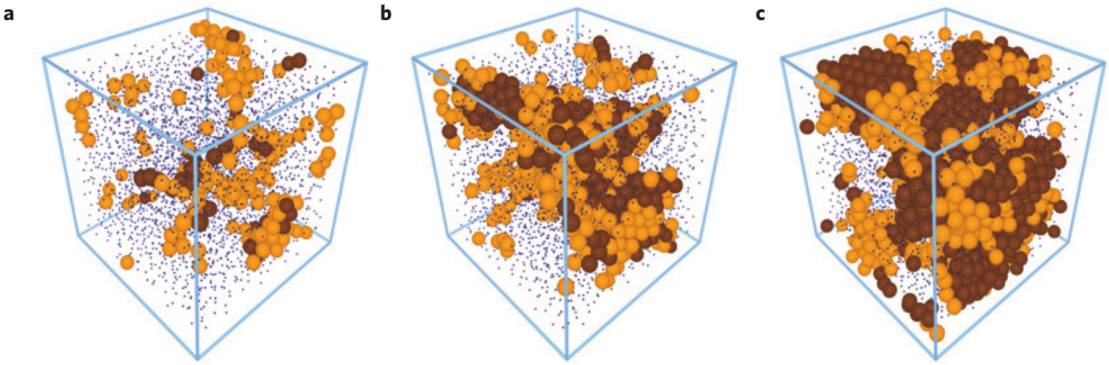

Figure 1.5: *Crystallization via a precursor stage. Blue dots represent liquid particles, orange spheres relatively ordered precursor particles and brown spheres nucleus particles. (a) Beginning of nucleation from precursors. (b) Subsequent development of precursors around critical nuclei. (c) Subsequent development of precursors around nuclei of post-critical sizes. Figure reproduced from Ref. [190].*

ferences between protein and colloidal systems, these observations are important and provide new insights for our understanding of crystallization processes.

Nonclassical crystallization processes have recently been discussed for protein systems. This is the topic of the next section. In contrast to simple systems, protein solutions cannot be sufficiently well described by easy theories. Protein–protein interactions in a salt solution are the result of a subtle balance between various forces such as Coulomb interactions, van der Waals forces, and affinity of aqueous anions for hydrophobic patches on the protein surface [104,105]. The existence of bonding patches implies that the proper spacial orientation of the crystalline connection also has to be considered [127].

1.2.3 Experimental Studies of Nonclassical Protein Crystallization

In recent years, several important experimental studies of nonclassical protein crystallization have been performed. In 1994, George and Wilson found by experimental observations that the second virial coefficient A_2 plays an important role for the crystallization of proteins [59]. Both repulsion and too strong attraction hinder the formation of crystals. The ideal conditions are given only within a narrow window of A_2 corresponding to weak attraction, the so-called crystallization slot. Methods frequently used to determine these interactions are static and dynamic light scattering (SLS, DLS), small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) [21–24,59,128,206]. Protein interactions in solution can be altered by the addition of ions. The strength of this influence depends on the number and valency of the ions.

The first prediction of a two-step pathway of protein crystallization by ten Wolde[196] was followed by experimental observations. Galkin and Vekilov studied the kinetics of the homogeneous nucleation of lysozyme and found that the nucleation rate is beyond the predictions of the CNT. It has a maximum in the vicinity of the LLPS boundary hidden below the solubility line in the phase diagram of lysozyme solution[53,54]. In Gibbs' original considerations, the initial and the final phase only differ by one order parameter, namely density. In the nucleation of crystals from solutions, however, they differ in at least two order parameters such as density and structure[202]. In the modeled systems, these two parameters occur simultaneously in off-critical compositions, similar to the classical setting. Close to the critical point of LLPS, they fluctuate independently from each other[202]. A crystal nucleates if a structure fluctuation occurs in a (short lived and due to a density fluctuation) region of locally higher density[202]. Other studies pointed out that a maximum in crystal nucleation rate can be found close to the liquid–liquid coexistence line, where the strongest density fluctuations occur[202].

This model suggests that the nucleation rate is higher in the protein-rich phase after LLPS because of the higher protein concentration[196]. Moreover, the surface energy at the interface between the dilute solution and the crystal is higher than at the interface between dense liquid and crystal[222]. This could be confirmed experimentally in some cases (e.g. for glucose isomerase[208] or hemoglobin[52]). However, experimental observations also suggest that protein crystals in several systems nucleate predominantly in the dilute phase or at the interface of both phases, growing into the dilute phase[30,100,148]. The dense phase merely acts as a protein reservoir in this case[100]. A possible explanation for this different mechanism could be that the stronger protein–protein interactions in the dense phase could hinder the reorientation and adjustment of conformation of the proteins towards a crystal[100].

As stated before, it is difficult to follow the nucleation process of proteins directly. Indirect experimental studies were performed applying quasi-elastic light scattering, dynamic light scattering, X-ray and neutron scattering[70]. Extracting the distribution of particle sizes from the scattering data can provide information about the nucleation kinetics. However, it depends on various assumptions concerning e.g. the particle shape[70]. It is also not clear if the observations only involve homogeneous nucleation or heterogeneous nucleation as well[70].

1.2.4 Theoretical Understanding and Simulations of Nonclassical Protein Crystallization

The topic of this section are theoretical studies and simulations of nonclassical protein crystallization processes. After George and Wilson described the crystallization slot, Rosenbaum *et al.* established a link between their work and observations that interactions between colloids can be tuned by adding polymers with varying radii of gyration[71,153,154,197]. They showed that the temperature range in which

a large number of globular proteins crystallizes, maps on a narrow range around the metastable critical point in the phase diagram calculated for colloids with short-ranged attractions. At too high temperatures, no crystallization is observed at all and at too low temperatures, the formation of aggregates dominates [71,197]. A similar crystallization window exists for colloids [197].

In 1997, numerical simulations performed by ten Wolde and Frenkel showed that the presence of a fluid–fluid critical point changes the nucleation pathway in homogeneous crystal nucleation dramatically [196]: the Gibbs free energy barrier, which is the rate-limiting step in crystal nucleation, was calculated for a system of spherical particles with a protein-like phase diagram with short-range attractions. Clusters consisting of multiple particles could be classified into belonging to a liquid or a solid phase, depending on their local ordering. Furthermore, a local density could be assigned to each cluster. It was found that well below the critical temperature, an increase in crystallinity (number of solid-like particles belonging to a nucleus) proportional to the increase in density leads to the lowest free-energy path to a critical crystal nucleus. This corresponds to the formation of a highly crystalline cluster directly from the liquid phase. However, at the critical temperature, the path of lowest free energy is found for the formation of a liquid-like droplet with increasing density, followed by a structural change towards higher crystallinity. The nucleation barrier at the critical temperature is also significantly lower compared to other temperatures. The order parameters structure and density therefore do not necessarily develop in a coupled fashion [196]. Thus, more than a century after Ostwald formulated his step rule on the basis of macroscopic studies, findings in protein systems suggest an equivalent step-rule on the microscopic level [197].

Theoretical considerations of Talanquer and Oxtoby [189] show that close to the metastable critical point, the crystal nucleation of globular proteins changes dramatically and nucleation rates are increased by orders of magnitude.

Having analyzed a large series of simulation results, Vliegenthart and Lekkerkerker predicted optimum conditions for protein crystallization in phase diagrams with a metastable gas–liquid region around the critical point and below [209]. The first step in the pathways suggested by ten Wolde and Frenkel, the local increase in density, can be due to critical concentration fluctuations or the formation of liquid droplets after a metastable LLPS below the critical point [196,209]. Predicting the critical temperature of isotropic potentials with short-range repulsion and long-range attraction, Vliegenthart and Lekkerkerker considered both cases and connected experimentally observed ideal conditions for protein crystallization (by George and Wilson, see below) to phase diagram regions. Around the critical point, they are found in different regions of the phase diagram than below the critical point, as shown in Figure 1.6 [209].

It was also indicated by numerical simulations that at conditions typically used for protein crystallization, small model protein clusters do not maintain a crystalline structure and protein crystal nucleation follows a two-step pathway [102].

Vekilov proposed the Gibbs free energy pathways shown in Figure 1.7b-c of two-

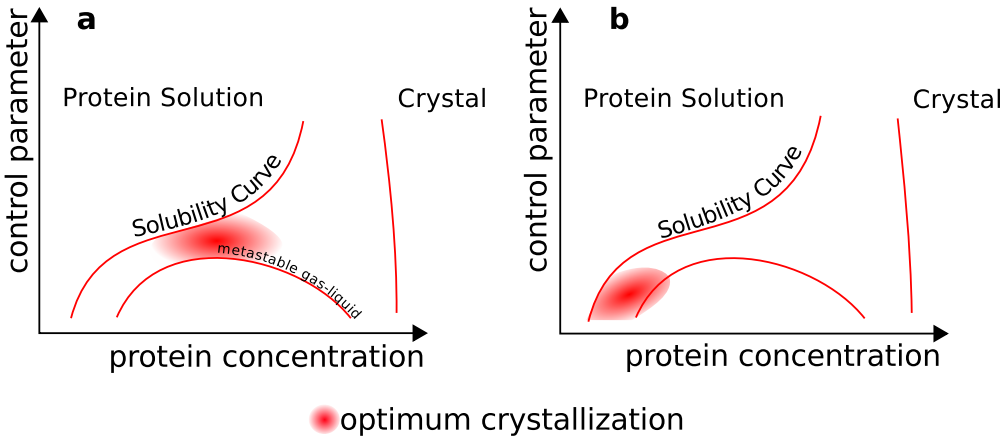

Figure 1.6: *Schematic phase diagrams indicating the optimum regions for crystallization (a) around the critical point. (b) below the critical point. Image adapted from Ref.* [209] *and modified.*

step crystallization processes. While in the classical pathway only one Gibbs free energy maximum is found (Figure 1.7a), a barrier that has to be overcome in order for nucleation to occur, the free energy landscapes of two-step pathways show an additional minimum corresponding to an intermediate phase. Close to the LLPS boundary, the intermediate phase is unstable and consists of mesoscopic clusters. In this case, the Gibbs free energy of the intermediate phase is higher than the free energy of the initial solution (Figure 1.7b). For a macroscopic dense liquid phase, the Gibbs free energy of the intermediate phase is lower than that of the initial solution and therefore more stable (Figure 1.7c). Both pathways, however, are metastable with respect to the crystalline phase [202,204]. Most probably, the precursor for crystallization in protein solutions is not a general property [203].

1.2.5 Open Questions

There are still some open questions concerning protein crystals and the details of their formation: obtaining good crystals that diffract in high resolution is still a bottleneck and can lead to significant delays in research undertakings [162].

It is difficult to disentangle crystal nucleation and growth, which is why the term "nonclassical crystallization" is often used instead of "nonclassical nucleation" [171]. Even "two-step nucleation" can be used for different scenarios of crystal nucleation from a precursor. One possible realization are two separate nucleation events from fluid to intermediate and from intermediate to the crystalline state. Another pathway referred to as two-step nucleation is the nucleation promoted by density fluctuations close to the LLPS spinodal or critical point, as originally proposed by ten Wolde and Frenkel [196]. These alternative pathways are energetically favored and therefore can enhance the probability of nucleation events.

How exactly a nucleation event starts is still controversial. The different possible

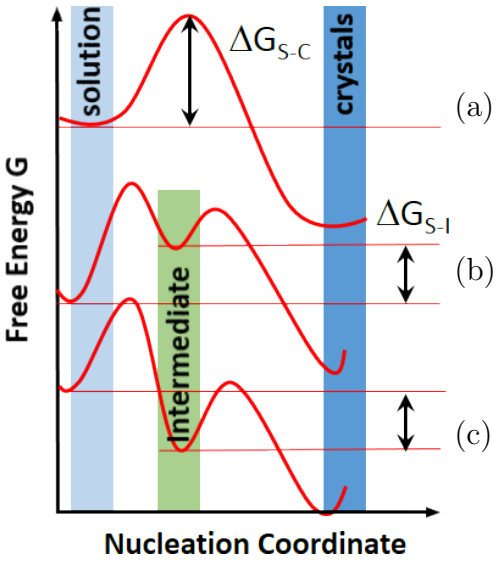

Figure 1.7: *(Gibbs) free energy pathways of crystallization processes. (a) Classical, one-step pathway with only one free energy barrier. (b) Two-step pathway with an intermediate phase metastable to both the initial fluid and the final crystalline state. (c) Two-step pathway with an intermediate phase stable with respect to the initial and metastable with respect to the final state. Image adapted from Ref. [202,222], above shown modified version is also found in publication A.*

mechanisms of nucleation and crystal growth are still not completely understood. Progress has been made in observing the nucleation process in real space using atomic force microscopy [94,111,119,181,201]. However, the characterization in ensemble average of such an early stage of nucleation in bulk is still challenging.

1.3 Ion Induced Phase Behavior in Protein Systems

Protein-protein interactions in a specific salt solution are the result of a subtle balance between different forces. A combination of long-range repulsion and short-range attraction leads to the formation of small equilibrium clusters [184]. At high ionic strength, salt screens the electrostatic repulsion and the attractive contributions dominate (increasingly with decreasing temperature). This can lead to LLPS and related phenomena [184]. These are also crucial for the crystallization of proteins and many aggregation-related diseases, as discussed in other sections of this thesis.

Influencing phase behavior in protein systems ranges back to the Hofmeister effect [78]. However, the focus of this section will be on phase behavior associated with reentrant condensation (RC) of negatively charged globular proteins.

1.3.1 Reentrant Condensation

RC is a concept known from systems such as DNA[133,134] or charged colloids[8]. The net charge of the respective particles changes sign with increasing concentration of a charged additive. At an intermediate additive concentration, the colloids, DNA or other polyelectrolytes are neutral and condensate. This phenomenon has recently also been found for negatively charged globular proteins in aqueous solution, triggered by the addition of trivalent salts such as YCl_3[220,226]. The diagram shown in Figure 1.8 illustrates this reentrant condensation in protein systems: it is divided into three regimes by two boundary salt and accordingly ion concentrations c^* and c^{**}. Below c^* (regime I), samples are clear, indicating stable solutions. At intermediate salt concentrations between c^* and c^{**} (regime II), samples become turbid due to aggregation and material falls out of solution. If c^{**} is exceeded, the condensed proteins redissolve and a one-phase state is reached again, the system "re-enters" (regime III). This can be explained by an initial increase in attraction between the negatively charged proteins due to cations binding to the initially negatively charged protein surface[220,226]. Beyond c^{**}, the strength of attraction decreases again and eventually, the positive charge is strong enough to lead to redissolution due to electrostatic repulsion. An effective charge inversion on the surface side chains is supported by zeta potential measurements and Monte Carlo simulations[220].

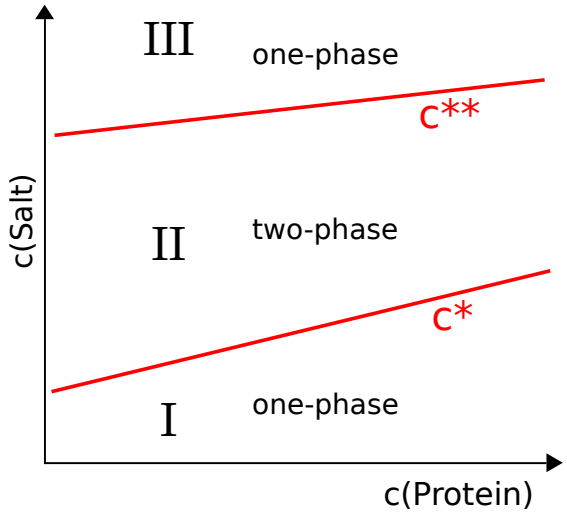

Figure 1.8: *Schematic reentrant condensation phase diagram. I, II, and III mark the three regimes.*

No condensation could be induced using the monovalent salt NaCl[220], nor by the large complex $SpeCl_4$, that is often successfully used for the condensation of DNA[226]. The latter is possibly due to the rather small charge density of $SpeCl_4$. Likewise, no RC could be observed if multivalent salts (including $SpeCl_4$) were added to solutions of the positively charged globular protein lysozyme[226].

The Derjaguin-Landau-Verwey-Overbeek (DLVO) theory is highly valuable in the colloidal context and also has long been used for the description of the basic phase behavior and particle–particle interaction in protein solutions. However, RC cannot be described sufficiently by DLVO because of the complex geometry and charge distribution[220]. On the other hand, the Debye-Hückel theory and related concepts work well for the complex geometry but cannot be applied for systems with strongly charged ions, where ion–ion correlations have to be considered and mean-field approaches therefore fail[220]. Likewise, theoretical approaches employed for colloids and DNA usually fail in this case due to the different shape and the more complex, irregular surface charge pattern[220].

Further, pH effects due to hydrolysis of metal ions should be taken into account. RC phase diagrams for highly acidic salts such as $AlCl_3$ or $FeCl_3$ show a much more narrow second regime compared with the rather neutral salts YCl_3 and lanthanum chloride ($LaCl_3$). This behavior can be explained by the interaction of pH effects and the binding of multivalent cations[151]. The results of Roosen-Runge *et al.* suggest that a combination of pH variation and multivalent counterions provides control over attraction and repulsion between globular proteins[151]. The addition of monovalent salt shifts c^* and c^{**} towards higher trivalent salt concentration[86]. This is reflected in the protein interactions as shown by SAXS and reproduced by simulation. The Coulomb repulsion is varied by the concentration of trivalent salt and screened by total ionic strength including the monovalent salt, while an attractive interaction is established by the multivalent counterions via bridging between proteins[86]. These results suggest that the interactions in a protein system showing RC can be tailored by salt effects[86]. The fact that nearly half of all protein families has a majority of acidic residues and thus are negatively charged at neutral pH gave rise to the hope that the phenomenon of RC could be observable in a large number of proteins[226].

1.3.2 Liquid–Liquid Phase Separation

Liquid–liquid phase separation (LLPS) is a phenomenon observed in a multitude of systems including protein solutions, where it was first found in the eye lenses of calf, rat and human[193,194], followed by a discovery in an aqueous lysozyme-salt solution[81]. In Subsection 1.1.2, the general principle was introduced. In this section, the focus is on LLPS in protein systems caused by multivalent metal ions.

In protein systems, LLPS is — caused by the relatively large size of proteins and therefore short-ranged attractions — metastable with respect to the crystalline phase, whereas a stable liquid phase exists for simple fluids[102]. The attractive forces between molecules that lead to a transition into a protein-rich and a protein-poor fluid also can enable crystallization[171]. Due to the missing energy barrier and hydrodynamic effects accelerating phase separation but not crystallization, LLPS through spinodal decomposition is significantly faster than crystallization and can affect its kinetic pathway[191].

In recent years, it has been observed that a liquid–liquid phase separation re-

gion often is found in a closed region of the second regime in a RC phase diagram. The system of HSA with YCl_3 was studied more in detail. In three dimensions (temperature, salt concentration and protein concentration), the LLPS region has the shape of an elliptic paraboloid. Under the microscope, denser droplets within the more dilute surrounding could be observed after a LLPS. It was confirmed by SAXS that the interactions are short-ranged attractive[224]. The isothermal binodal was determined and the reduced second virial coefficient was used in order to describe the interactions close to and far away from the binodal. It is negative (meaning attractive interactions) in both cases but increases approaching the critical point, in accordance with theoretical predictions for colloid systems[216].

In protein systems as well as in atomic or colloidal systems, an upper critical solution temperature (UCST) is usually observed where LLPS occurs upon cooling, in accordance to the explanation in Subsection 1.1.2. However, there are also systems with a retrograde temperature dependence, where LLPS occurs at high temperatures, e.g. several variants of hemoglobin[52,203]. This lower critical solution temperature (LCST) behavior also was observed in other systems such as polymers[12,130,163], polyethylene oxide – water systems[110] or aqueous solutions of lutidine[103] or nicotine[36]. In reentrant condensation systems, LCST behavior was found recently in BSA (and HSA) with YCl_3[25]. This cannot be explained by attractive forces only. Instead, an entropic contribution has to be taken into account. By increasing volume fraction of D_2O instead of H_2O as a solvent, the region of phase separation shrinks, until it has completely disappeared in pure D_2O. Isothermal titration calorimetry measurements show that the binding of Y^{3+} ions to a protein is entropy-driven. The LCST behavior is explained by the entropy gain due to released hydrated water molecules. At even higher temperatures, the LLPS region is expected to close again[25].

1.3.3 Ion Induced Crystallization of Proteins

While crystallization is known since antiquity, the first crystallization of a protein, hemoglobin, was reported in 1840[62]. Under certain conditions, CNT has been applied successfully for many systems, including protein solutions[180]. However, in various studies of the crystallization of proteins, features beyond this simple view have been discovered in the early stage of nucleation[171]. Compared to the crystallization of small molecules, protein crystallization requires unusually high supersaturation and occurs much more slowly[127]. A general difference arises from the relatively large size of proteins: these systems often show an attractive interaction short-ranged compared to the particle size (which also applies to colloids). Furthermore, the interactions between proteins are anisotropic due to their irregular shape and inhomogeneous charge pattern. Therefore, isotropic models fail to describe complex properties of proteins such as their nucleation behavior[70]. Patchy models are to some degree more suitable to protein systems and have therefore been used repeatedly in recent years[51,66,152,213,214].

Ideal crystallization conditions are found in a narrow range of weak attrac-

tions, the crystallization-slot[59]. In a repulsive system, the proteins stay in solution while too strong attractive interactions lead to stable amorphous aggregates. Recent results indicate that sign and strength of protein–protein interactions in a large number of systems can be modulated towards crystallization by RC[86,151,220,226,227]. The physical mechanism of RC has been demonstrated to be charge inversion[151,226]. The first regime at low salt concentrations is dominated by electrostatic repulsion. In the second regime, at intermediate salt concentrations, attractive forces become predominant which leads to the condensation of protein molecules and the formation of big clusters. In the third regime at high salt concentrations, proteins repel each other again. Specific binding of the cations to surface-exposed side chains leads to the formation of ion bridges between protein molecules[227]. The cations enable the formation of the crystal lattice and in addition can be used to solve the phase problem[227]. Thus, the effective protein-protein interactions can be modulated by variation of the salt concentration. A similar method was successfully applied for basic proteins and multivalent anions: in a system of a tyrosinase isoform from *Agaricus bisporus*, Mauracher *et al.* successfully added $MgCl_2$ or $Na_6[TeW_6O_{24}]\cdot 22H_2O$ (TEW) in order to induce crystallization[116]. Hen egg-white lysozyme was likewise crystallized in the presence of TEW. The negatively charged TEW molecules are part of the crystal lattice and bind to positively charged protein sites[20].

1.4 Outline

This thesis is divided into three parts. Part I introduces a motivation and fundamentals for the understanding of this work as well as the used materials and methods. Part II comprises the results and is predominantly based on publications in scientific journals (labeled as publications A-D). Part III summarizes the results and draws conclusions. It subsequently gives an outlook on probable developments and possible future work. The composition of *Results & Discussion* (Part II) is listed in the following.

- [A] A. Sauter, M. Oelker, G. Zocher, F. Zhang, T. Stehle, and F. Schreiber.
  Nonclassical pathways of protein crystallization in the presence of multivalent metal ions.
  Cryst. Growth Des. **14** (2014), 6357–6366.

- [B] A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, R.M.J. Jacobs, and F. Schreiber.
  Real-time observation of nonclassical protein crystallization kinetics.
  J. Am. Chem. Soc. **137** (2015), 1485–1491.

- [C] A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, A. Feoktystov, R.M.J. Jacobs, and F. Schreiber.
  On the question of two-step nucleation in protein crystallization.
  Faraday Discuss. **179** (2015), 41–58.

- [D] A. Sauter, F. Zhang, N. K. Szekely, V. Pipich, M. Sztucki and F. Schreiber.
  Structural evolution of metastable protein aggregates in the presence of trivalent salt studied by (V)SANS and SAXS.
  In preparation.

- *Complementary study: on the universality of inducing protein crystallization by RC.*
  - Toolbox for tuning phase behavior in protein solutions using trivalent salts
  - EpzP - a “real-world system”

Chapter 2

Experimental

2.1 Materials and Sample Preparation

2.1.1 Proteins

The protein that was mainly used for the work presented in this thesis is β-lactoglobulin (BLG) since it proved to be a good model system for the temperature-dependence of the upper reentrant condensation boundary and crystallization induced by multivalent metal salts[221,227]. An illustration is found in Figure 2.1. The protein belongs to the group of lipocalins, small extracellular proteins, and is the main protein of whey in the milk of many mammals[49]. Its physiological function — if other than nutritional — remains unclear[162]. Transport and immunoregulatory functions have been discussed[49]. Around ten genetic variants have been identified, but A and B are by far the most common ones[162]. For this work, BLG from bovine milk, namely a mixture of the genetic variants A and B, that differ at two of 162 amino acids in total (exchange of Asp64Gly and Val118Ala)[162,207], was purchased from Sigma-Aldrich (Taufkirchen, Germany, *product no. L*3908, 90% purity). A monomer has a molecular mass of about 18.3 kDa. At room temperature and a pH between 3.5 and 7.5, BLG in aqueous solution has a negative surface charge (isoelectric point of 5.2) and is found predominantly in dimer state[44,207].

Additional proteins were used for the work presented in Chapter 7. The proteins used in Section 7.1 were purchased from Sigma-Aldrich, too: serum albumins are prevalent transport proteins in the blood plasma of mammals with a molar mass of about 66 kDa[28,109]. Bovine serum albumin (BSA, 99%, A3059) and human serum albumin (HSA, 97-99%, A9511), both with an isoelectric point (pI) of 4.7[178], were used in this study. The digestion protein pepsin from porcine gastric mucosa (3200-4500 units/mg, P6887) has a molecular weight of 34.62 kDa and a pI between 2.2 and 2.8[178]. Ovalbumin (OVA, 98%, A5503) from chicken egg white has a molecular mass of 44 kDa and a pI of 4.54[178]. The enzyme EpzP (Chapter 7.2) was recently isolated and characterized biochemically. It is a globular, monomeric protein with a very rigid barrel structure (diameter ≈ 45 Å, height ≈ 32 Å) and with 302 amino acids, it has a molecular mass of 33.2 kDa. The theoretical pI is 5.3 as a result of 40 Asp/Glu residues compared to 32 Lys/Arg residues. The protein plays an important role in the biosynthesis of endophenazines[229].

In the case of most proteins, protein stock solutions were prepared by mixing degassed MilliQ (18.2 MΩ, Merck Millipore, Darmstadt, Germany) water and pro-

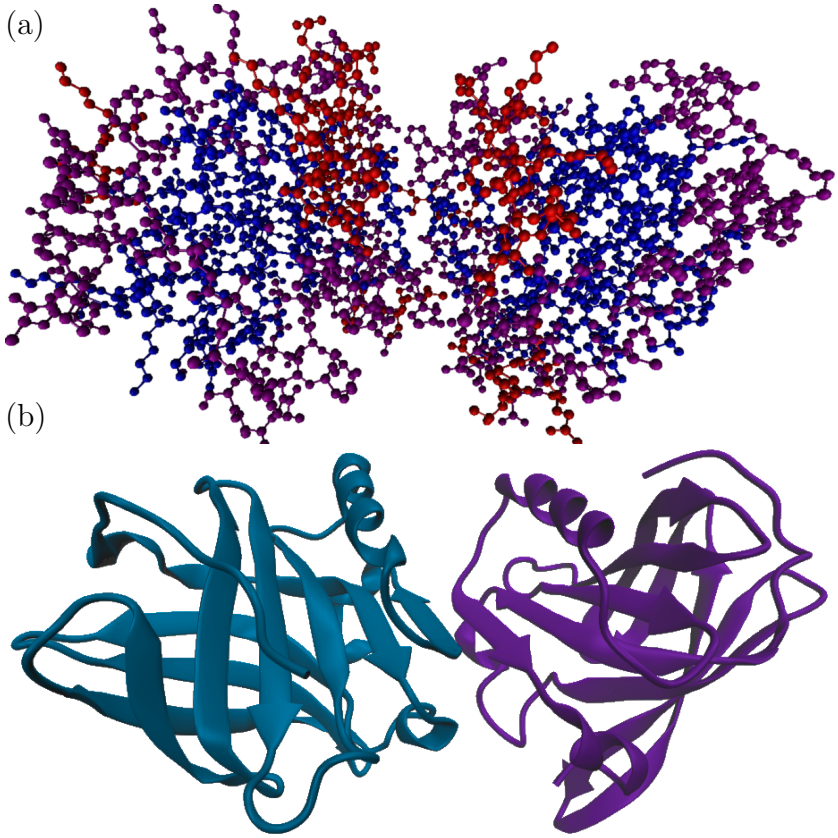

Figure 2.1: *Asymmetric unit of BLG (dimer) from the pdb entry 3ph5*[227]*. The pictures were created with Visual Molecular Dynamics using the pdb data and modified afterwards. (a) Representation of all atoms (except hydrogen) as spheres. α-helices are shown in red, β-sheets in blue and other structures in purple. (b) "Cartoon" of the secondary structures. The two monomers are shown in different colors.*

tein powder and keeping the mixture at 5°C for one day for complete dissolving. Exceptions were OVA and EpzP: OVA solutions were purified by size exclusion chromatography. Typically, this protein was purified employing a flow rate of 1 mL/min and a buffer consisting in 50 mM HEPES (pH 7.5) and 150 mM NaCl. Fractions corresponding to the Ovalbumin monomer were pooled and buffer was exchanged with degassed MilliQ water using 10 kDa MWCO centrifugal concentrators and used as stock solutions. EpzP was produced in buffer (20 mM HEPES at a pH of 7.5, 150 mM NaCl) which was — if pure water as a solvent was desired — exchanged afterwards by degassed MilliQ water using centrifugal concentrators. The real protein concentration of the stock solutions was measured subsequently by ultraviolet-visible (UV-Vis) absorption spectroscopy (see below). If not in use, protein stock solutions were stored at 5°C.

2.1.2 Salts

The multivalent salts mainly used in this work are zinc chloride ($ZnCl_2$), cadmium chloride ($CdCl_2$), and YCl_3. They were purchased from Sigma Aldrich with guaranteed purities of 97% for $ZnCl_2$ (96468), 99.99% for $CdCl_2$ (202908), and 99.99% for YCl_3 (451363). For the work presented in Section 7.1, the following additional salts, also purchased from Sigma Aldrich, were used: $LaCl_3$ (449830), cerium chloride ($CeCl_3$) (450731), gadolinium chloride ($GdCl_3$) (439770), yttrium nitrate ($Y(NO_3)_3$) (237957, hexahydrate), yttrium sulfate ($Y_2(SO_4)_3$) (326070, octahydrate), ytterbium chloride ($YbCl_3$) (450073), ytterbium nitrate ($Yb(NO_3)_3$) (217220, pentahydrate) and ytterbium sulfate ($Yb_2(SO_4)_3$) (575089). All of the cations in these salts belong to the rare earth elements, which consist of Sc, Y and the lanthanoids (La and the fourteen following elements)[35]. Salt stock solutions were prepared by giving the desired mass of salt on the basis of their molecular weight into a volumetric flask which was subsequently filled with degassed MilliQ water to dissolve the salt.

2.1.3 Sample Preparation and Survey of Solution Properties

For sample preparation, deionized (18.2 MΩ) and degassed MilliQ water (H_2O) was used. An exception were samples for neutron scattering experiments and infrared spectroscopy, which were prepared in heavy water (D_2O) with a purity of 99.9 atom % deuterium. To prevent the exchange with H_2O from the air, paraffin strips were wrapped around vials with D_2O and D_2O-based solutions. This also prevents CO_2 from dissolving in H_2O and D_2O solutions. All samples were prepared by mixing protein stock solution, pure water and the respective salt stock solution in the needed fractions. The used pipettes (purchased from Eppendorf, Hamburg, Germany) have a a relative error between 0.2 and 1% and a systematical error between 0.6 and 3%[46]. The sample was homogenized by careful shaking or by stirring with a pipette-tip. The room temperature was 21 °C, mediated by air conditioning. If other or more exact temperature settings were needed, a water bath was applied (±0.01 °C).

Ultraviolet-Visible Absorption Spectroscopy

Due to non-protein impurities and the own volume of the proteins, the real concentration differs from the prepared one (measured mass of protein divided by added volume of water). Therefore, the real protein concentration of every stock solution and, in some cases, of samples was measured by UV-Vis absorption spectroscopy. This was accomplished by means of a Cary 50 UV-Vis spectrophotometer from Varian Inc. Santa Clara, USA with the software Cary WinUV.

Amino acids absorb ultraviolet light. Tyrosine, tryptophan, and, to a much lesser extent, phenylalanine, are known to cause a characteristic strong absorbance

of most proteins at a wavelength of around 280 nm which can be used to determine the concentration of a protein if the ratio of the aforesaid amino acids (given by the so-called extinction coefficient) is known [95]. The fraction of the incident light that is absorbed is proportional to the sample thickness d, the concentration c and the extinction coefficient ϵ, as described by the Lambert-Beer law:

$$log\frac{I_0}{I} = \epsilon cd \tag{2.1}$$

where I_0 is the intensity of incoming and I the intensity of transmitted light. $log\frac{I_0}{I}$ is also called "absorbance" [95]. ϵ is often given in units of $l/(mol \cdot cm)$ (molar extinction coefficient, for concentrations in $M = mol/l$). In this work, the specific extinction coefficient in units of $l \cdot g^{-1} \cdot cm^{-1}$ and protein concentrations in $mg/mL = g/l$ are used.

The absorbance of the same cuvette filled with degassed MilliQ water has been measured as a background and subtracted from the measurement of the protein solution.

Size Exclusion Chromatography

Size exclusion chromatography (SEC), also called gel filtration, is a common technique for the purification of proteins or for analysis of the constituents of a protein solution. A column is filled with a porous material and a buffered, protein-containing solution percolates through it. Larger proteins migrate faster because they cannot enter the cavities of the porous beads of cross-linked polymer inside the column. Taking advantage of this mechanism, the proteins are separated into fractions based on their size [95]. Figure 2.2 yields an illustration. The amount of protein in each fraction is determined by UV-Vis absorption at λ=280 nm. Other types of column chromatography include separation according to the charge, binding affinity and other properties of the proteins [95].

Survey of pH Values

A Seven Easy pH instrument from Mettler Toledo, Columbus, USA was used to monitor the change of the pH of protein solutions caused by salt addition. For every pH value, the protein acquires a different surface charge [127]. If the pH fell below the pI, this would invert the charge of the proteins used in this work which are negatively charged at neutral pH. The pH changes caused by the addition of salts in samples prepared for the work presented in this thesis were not strong enough to reach or cross the pI. Thus, cation binding must be the main driving force of the observed charge inversion [151,226]. Attention should be paid also to the pH range in which the respective protein is stable and in its native conformation. Regular pH measurements ensured that this range was not exceeded during our experiments.

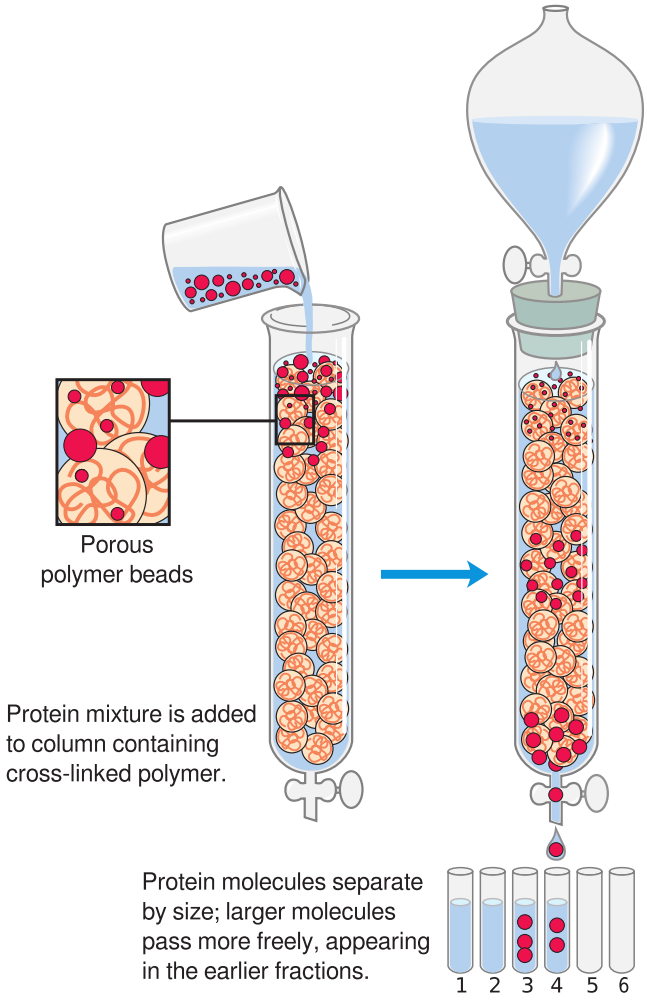

Figure 2.2: *Illustration of the size exclusion chromatography method. Larger proteins migrate faster through the column matrix with pores of selected size. Smaller proteins enter the pores and are slower because they take a longer path through the column. Image adapted from Ref.* [95].

2.2 Survey of Protein Stability

In order to ensure that no significant structural changes occur in the proteins under the experimental conditions, Fourier transform infrared (FTIR) and circular dichroism (CD) measurements were performed.

FTIR allows to monitor the stability of the protein secondary structure in D_2O. Infrared spectra result from the excitation of vibrational energy states in molecules due to coupling of the electric field vector with the dipole moment of the molecules. Each mode translates into a specific frequency[14,69]. From the vibrational spectrum, information on the chemical structure, redox state, bond lengths and strength, bond angles and conformation and more can be derived. In the case of proteins, absorption of infrared light by amino acid side chains

provides some information when protein reactions are investigated, but usually the amide modes are investigated which are caused by vibrations of the peptide groups[14]. In this thesis, the amide I and amide II bands are surveyed upon temperature change and the addition of multivalent ions. Amide I is mainly due to the C=O stretching vibration, while amide II is the out-of-phase combination of C-N stretching vibration and N-H inplane bend[14]. Both are sensitive to secondary structure (especially amide I at $\sim$1650 cm^{-1})[14]. Using a Michelson interferometer, a collimated beam of radiation is split into two parts which are recombined and interfere. By varying the path length difference with constant velocity, the wavelength with maximum constructive interference is "selected" from the polychromatic source[69]. As an instrument, the IFS 48 from Bruker (Billerica, USA) was used. Protein solutions for FTIR measurements were prepared in D_2O because the absorption bands of H_2O and proteins interfere. All spectra were corrected by a background measurement of pure D_2O.

CD is a technique that allows to investigate the secondary structure of proteins taking advantage of the circular polarization[70]. Plane polarized light can be understood as a combination of two circularly polarized components of equal magnitude and different rotation direction. If these components are not absorbed to equal extends by a sample, elliptically polarized light is detected as an output signal. This is observed when the sample is optically active because it has a chiral structure. This can be e.g. due to a carbon atom with four different substituents, due to placement in an asymmetric environment (three-dimensional structure of the molecule) or because it is covalently bound to the chiral molecule center[89]. In proteins, the CD signal is caused by the peptide bond, aromatic side chains and disulphide bonds as well as by non-protein cofactors and ligands that acquire chirality when bound in the asymmetric environment provided by the protein. Information that can be obtained from protein samples by means of CD include the secondary structure composition, a certain fingerprint of the tertiary structure, and conformational changes such as protein folding[89]. For the CD measurements performed for this thesis, a J-720 spectrophotometer from Jasco Inc. (Easton, USA) was used.

2.3 Zeta Potential Determination

The charged surface of a protein in solution is surrounded by a distribution of ions. The rather rigid inner shell formed by these ions is called *Stern layer*, the ions further away from the protein form the *diffuse layer*[39]. The boundary between the hydrodynamically mobile and immobile fluid is called slip plane or slipping plane[39]. The ions within this boundary move with the particle. The potential between the solution and the moving particle is called zeta potential[112]. An illustration is found in Figure 2.3. The isoelectric point pI is the point at which the zeta potential is equal to 0[39].

The measurements were performed with a Zetasizer Nano ZS from Malvern

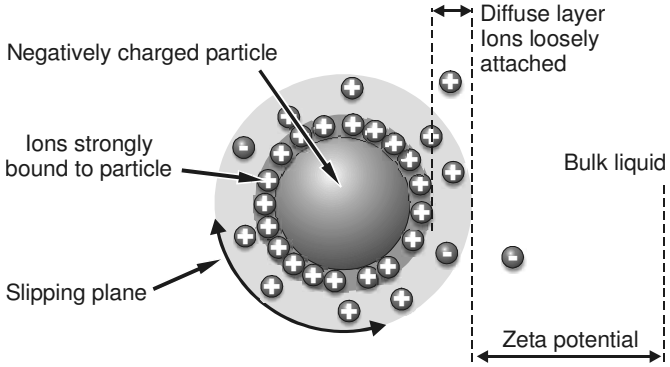

Figure 2.3: *Illustration of the zeta potential. Image reproduced from Ref.* [112].

Instruments Ltd, Worcestershire, UK. An alternating electrical field was applied to a U-shaped capillary cell. Using laser Doppler electrophoresis with a 633 nm laser at a fixed angle of 173°, the velocity of particles in the electrical field is measured from which the zeta potential is calculated. All samples were filtered using a 100 nm filter before the measurement. Zeta potential measurements were performed to confirm the charge inversion in protein systems upon adding an appropriate amount of salt.

2.4 Optical Microscopy and Crystallization Methods

The crystallization of samples was followed by means of optical microscopy. The microscope mainly used was an AxioScope A1 from Carl Zeiss Microscopy (Oberkochen, Germany) which allows automatized time-dependent picture series by the software Zen 2012. Other employed microscopes are an Axioskop 40 from Carl Zeiss Microscopy in a 277 K room and a DME from Leica Microsystem GmbH (Wetzlar, Germany) in a 293 K room with attached cameras.

Prior to the crystallization experiments, protein stock solutions were filtered using a 100 nm syringe filter in order to remove dust. The protein concentration was determined afterwards by UV-Vis absorption spectroscopy. Sample conditions were tested for crystallization using (micro-)batch setups in which drops of the sample are prevented from evaporation [70]. This is also the underlying method for most experiments followed by optical microscopy. For means of comparison, experiments at several conditions were additionally performed using the vapor diffusion hanging drop method [62]. In vapor diffusion methods, a drop of the solution of protein, buffer and other ingredients such as salts is placed close to a sealed reservoir with the same solution without protein. The two solutions are not in contact with each other. The concentration difference drives the system towards equilibrium concentration: water from the drop evaporates and is transferred to the reservoir [70].

For the quantitative analysis of microscopy pictures, samples were prepared in a micro-batch setup between two hydrophobically coated glass slides sealed by silicone. The small sample thickness of approximately 250-300 µm minimizes the effect of crystallization in different layers and keeps the number of crystals manageable. The hydrophobic surfaces avoid wetting and heterogeneous nucleation. Indeed, a significantly larger number of crystals was observed using glass plates without hydrophobic coating. The number and visible area of crystals was evaluated by the open source program *ImageJ* [167]. Although a python program for the automatized recognition of crystals (based on a script from F. Zanini) is being developed, the crystals evaluated for the publications included in this thesis were marked by hand since this method is still clearly more accurate. The average length of the crystals in an image was estimated as $L \approx \sqrt{A/N}$ with the number of crystals N and the visible area of crystals A. L and N were plotted as a function of time (t). The nucleation rates can be estimated from the slope of a linear fit of $N(t)$, whereupon possible crystallites with a size below the observable length scale that merge (Ostwald ripening) or do not grow large enough to be visible are neglected.

2.5 Small Angle Scattering

Small angle scattering is widely used to determine protein–protein interactions in aqueous solutions. For the studies presented in this thesis, both small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) were applied. Although there are naturally physical differences, the mathematical formalism and most data analysis are similar for X-ray and neutron scattering [188]. In the following, the basic concept is summarized. The basic setup of a scattering experiment is shown in Figure 2.4. A beam of radiation or particles, respectively, is deviated from its former trajectory by inhomogeneities in the sample it passes. Since the distances of the setup (source–sample, sample–detector) are much larger

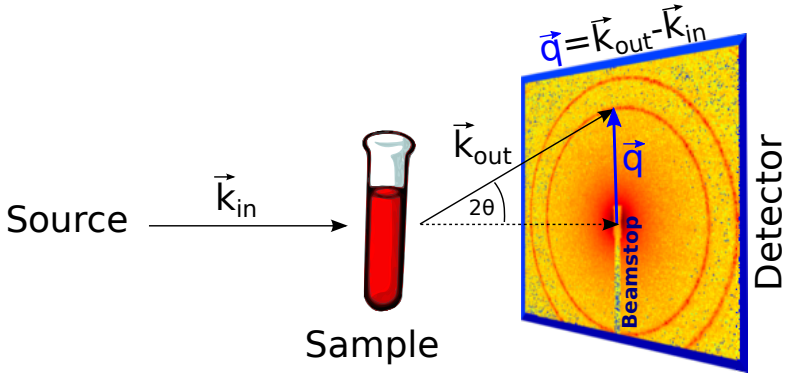

Figure 2.4: *Basic setup of a small angle scattering experiment. The 2D-image is obtained from a SAXS measurement of 33 mg/mL BLG with 17 mM $CdCl_2$.*

than typical sizes of inhomogeneities in the sample that are still larger than the wavelength, the incoming wave vector $\vec{k}_{in}$ and the outgoing, scattered wave vector $\vec{k}_{out}$ can be considered as plane waves[188]. We will only describe elastic scattering, where energy and wavelength of incoming and scattered beam are the same[188] and which is isotropic and thus only depends on the momentum transfer $\vec{q}$ with

$$|\vec{q}| = |\vec{k}_{out} - \vec{k}_{in}| = \frac{4\pi sin(\theta)}{\lambda} \tag{2.2}$$

where 2θ is the scattering angle and λ the wavelength[70,188]. The strength of interaction between beam and sample is given by the differential scattering cross section $(d\sigma/d\Omega)$ defined as

$$\frac{d\sigma}{d\Omega} = \frac{I_{out}}{j\Delta\Omega} \tag{2.3}$$

with the flux j which is the number of photons (SAXS) or neutrons (SANS) in the incident beam passing through a unit area per second and the solid angle $\Delta\Omega$[6,188].

The experimentally accessible intensity I (that is directly related to the structure of the sample) scattered by a unit volume can be defined as

$$I = \frac{1}{V}\frac{d\sigma}{d\Omega} = \frac{\Delta j}{j}\frac{1}{\mathcal{T}d\Delta\Omega} \tag{2.4}$$

with Δj being the fraction of flux that is elastically scattered in the direction of the solid angle, the sample thickness d and its transmission $\mathcal{T}$[98]. The measured total scattering intensity of different types of non-interacting particles can be written as a linear combination of their individual scattering intensities[188]. The scattered radiation $I(\vec{q})$ depends, for a given beam and illuminated element j, on the scattering length b_j and the scattering length density ρ_j. The total scattering length density ρ, that is, the density of scatterers of type j, can be written as $\rho(\vec{r}) = \sum_j \rho_j(\vec{r})b_j$[98]. It has different origins in X-ray and in neutron scattering: X-rays interact with the electron shell and neutrons interact with the nuclei (and spins)[188]. For SAXS and SANS, the Born approximation can be applied, assuming that the interactions of the beam with one scatterer do not influence the scattering by others. Thus, the total scattering amplitude $(\mathcal{A})$ for the illuminated volume V is given by[98]

$$\mathcal{A}(\vec{q}) = \int_V \rho(\vec{r})e^{-i\vec{q}\vec{r}}d\vec{r} \tag{2.5}$$

and the measured intensity as

$$I(\vec{q}) = \frac{\mathcal{A}(\vec{q})\mathcal{A}^*(\vec{q})}{V} = \frac{1}{V}\int_V\int_V \rho(\vec{r})\rho(\vec{r'})e^{-i\vec{q}(\vec{r}-\vec{r'})}d\vec{r}d\vec{r'}. \tag{2.6}$$

In particulate systems, a form factor $P(\vec{q})$ and a structure factor $S(\vec{q})$ can be

defined. The intensity for a single particle is given by

$$I_{part}(\vec{q}) = \mathcal{A}_{part}(\vec{q})\mathcal{A}^*_{part}(\vec{q}) = V^2_{part}P(\vec{q}) \tag{2.7}$$

with the form factor

$$P(\vec{q}) = \frac{1}{V^2_{part}} \int\int \rho(\vec{r})\rho(\vec{r'})e^{-i\vec{q}(\vec{r}-\vec{r'})}d\vec{r}d\vec{r'} \tag{2.8}$$

of the particle[98]. For several diffusing particles that interact with each other, the scattering intensity varies and the time average is measured instead of Eq. 2.6, which is equivalent to the statistical average for ergodic samples. If the particles are spherical (or parallel in a nematic / smectic phase) with identical interactions, the expression further simplifies and the total scattering intensity can be written as

$$I(\vec{q}) = \Phi V_{part}P(\vec{q})S(\vec{q}) \tag{2.9}$$

with the volume fraction Φ and the structure factor

$$S(\vec{q}) = 1 + \frac{N-1}{V}\int_V g(\vec{r})e^{-i\vec{q}\vec{r}}d\vec{r}. \tag{2.10}$$

$g(\vec{r})$ is the pair correlation function of the centers of mass of the particles[98].

In order to describe the effective interactions, structure factors are calculated from interaction potentials. The potentials used for the structure factor fits in this work (after radial averaging of the intensity) were, depending on the phase diagram region, a screened Coulomb potential, a hard sphere or a sticky hard sphere potential. The hard sphere potential (with the hard sphere diameter σ) is defined as

$$U_{HS}(r) = \begin{cases} \infty & r \leq \sigma \\ 0 & r > \sigma. \end{cases} \tag{2.11}$$

The sticky hard sphere potential is defined as[15]

$$U_{SHS}(r) = \begin{cases} \infty & r < \sigma \\ -u & \sigma < r < \sigma + \Delta \\ 0 & r > \sigma + \Delta. \end{cases} \tag{2.12}$$

with the stickiness parameter

$$\tau = \frac{1}{12\epsilon_\Delta} exp(-u/kT) \tag{2.13}$$

and the perturbation parameter ϵ_Δ

$$\epsilon_\Delta = \Delta/(\sigma + \Delta) \tag{2.14}$$

Δ is the width of the square well, and u is its depth[90]. Finally, the screened Coulomb potential is defined as[73,77]

$$U_{SC}(r) = \begin{cases} \frac{Z^2 e^2}{\varepsilon(1+\kappa_D\sigma/2)^2} \frac{exp(-\kappa_D(r-\sigma))}{r} & r > \sigma \\ \infty & r \leq \sigma \end{cases} \tag{2.15}$$

with Z the protein surface charge, e the electronic charge, and ε the dielectric constant of the solvent. κ_D is the inverse of the Debye screening length, determined by the ionic strength of the solution and σ is the effective sphere diameter of the ellipsoid the axes a, b, b.

2.5.1 Peculiarities in SANS and SAXS

Whether SANS or SAXS are more suitable depends on the sample of interest. Some samples have more contrast for neutrons and others for X-rays, which is connected to differences in the scattering process[98]: In the case of neutrons, the scattering length b_j (with $d\sigma/d\Omega(\vec{q}) = b_j^2$ for a point scatterer) depends on the isotope of the sample and its spin and does not depend on the atomic number in a monotonic way[179]. For X-rays, b_j decreases monotonically with decreasing λ and increasing θ and is proportional to the atomic number, since the scattering depends on the number of orbital electrons and can therefore be described by the local electron density[98,179]. Near atomic binding energies, the atomic scattering factors are described by additional terms ("anomalous" scattering)[145].

By choosing solvents with appropriate electron density, the scattering of certain components of the sample can be masked out in SAXS experiments. However, for contrast variation studies, usually SANS is applied because of the enormous difference in the interaction of neutrons with hydrogen (^1H) and deuterium (^2H)[145]. In SAXS experiments, radiation damage can occur — especially to biological samples — due to the high flow of energy[219].

2.5.2 Sample Preparation and Data Analysis

For the work presented in this thesis, SAXS measurements were performed at the beamline ID2 at the ESRF (European Synchrotron Radiation Facility, Grenoble, France) using a flow-through capillary cell or, in the case of time-resolved measurements, a vertical quartz capillary. Exposure times were between 0.05 and 0.1 s. For time-dependent SAXS, the measurement started about 2-3 min after mixing. To ensure the consistency of the measurements, the sample was shifted vertically and data were collected from different positions. This cycle was repeated every 2-5 min. SANS measurements were performed in rectangular quartz

cells with a path-length of 1-2 mm. All SANS data presented in this thesis were collected at FRM2 (Garching, Germany) at the instruments KWS1, KWS2, and, for very small angle neutron scattering (VSANS), at KWS3. The two-dimensional intensity pattern was averaged azimuthally.

At low protein concentrations and with NaCl for screening, the form factor of BLG was determined experimentally. Naturally, it can be fitted very well by the crystal structure of the BLG dimer (PDB code of 1BEB[27]) using CRYSOL[186]. Spheres or ellipsoids, however, have also been shown to be good models for the form factor of BLG dimers or clusters. To describe the effective interactions, structure factors (using the average structure factor approximation from a monodisperse spherical system[31] with effective sphere diameters) derived from model potentials were used for data analysis. Data fitting for the determination of the interaction potential was performed using IGOR Pro with macros developed at NIST[90].

While a so-called Guinier plot gives the radius of gyration R_g of the particles and a Porod plot gives information about scattering inhomogeneities such as surface roughness through the Porod exponent[72], the Beaucage model is well suited for the analysis of small-angle scattering data from complex systems with multiple levels of related structural features (e.g. fractal and particulate systems). It yields multiple sets of radii of gyration and power law exponents from different structural levels by modeling the Guinier region and the Porod region[16,17,72]. It is characterized by three fitting parameters per structural level: the radius of gyration R_g, the Porod exponent d and the Guinier scaling factor G (from which the Porod scaling factor B can be derived)[72]. The scattering intensity can then be expressed as

$$I(q) = G \cdot exp\left(\frac{-q^2 R_g^2}{3}\right) + B \cdot \left[erf\left(\frac{qR_g}{\sqrt{6}}\right)^3 \cdot \frac{1}{q}\right]^d \tag{2.16}$$

for one Guinier and one Porod region, or, in the case of multiple levels, as[16]

$$\begin{aligned} I(q) \approx \sum_{i=1}^{n} \Bigg\{ & G_i \cdot exp\left(\frac{-q^2 R_{g(i)}^2}{3}\right) \\ & + B_i \cdot exp\left(\frac{-q^2 R_{g(i+1)}^2}{3}\right) \cdot \left[erf\left(\frac{qR_{g(i)}}{\sqrt{6}}\right)^3 \cdot \frac{1}{q}\right]^{d_i} \Bigg\}. \end{aligned} \tag{2.17}$$

For the analysis performed in the publications B and C, the degree of crystallinity was monitored. All curves were divided by the first one. Two Bragg peaks and a broad peak that we interpret as typical length within larger aggregates (and therefore as a measure for the amount of precursor phase) were fitted by scaled Gaussians. The areas under these Gaussian functions were plotted in dependence of time and further discussed.

2.6 Structure Determination by X-ray Diffraction

In 1934, X-ray crystallography entered biology by the first picture of a protein (pepsin) crystal[19,62]. Since then, protein crystallography has developed to the stage where the protein data bank (PDB) contains over one hundred thousand biological macromolecular structures[1].

Like SAXS, X-ray crystallography involves a sample placed into a collimated beam of X-rays and measuring the intensities of the scattered photons[145]. In solution scattering, the signal is averaged over all molecule orientations relative to each other as well as to the experimental setup and is therefore radially symmetric and continuous. Diffraction from a crystal, in contrast, results in a much larger signal of discrete diffraction maxima caused by the convolution of the lattice onto the continuous transform. In contrast to SAXS, crystallography allows the determination of structures at atomic resolution[145]. Single crystals exhibit a long-ranged translational order that leads to interference of diffracted X-rays and a characteristic spot pattern[34]. An example is shown in Figure 2.5. These spot positions allow the calculation of lattice plane distances and the orientation of the crystal lattice[34]. Polycrystalline aggregates, in contrast, lead to a ring-like pattern, known as powder diffraction, due to the random orientation of crystalline building units[34].

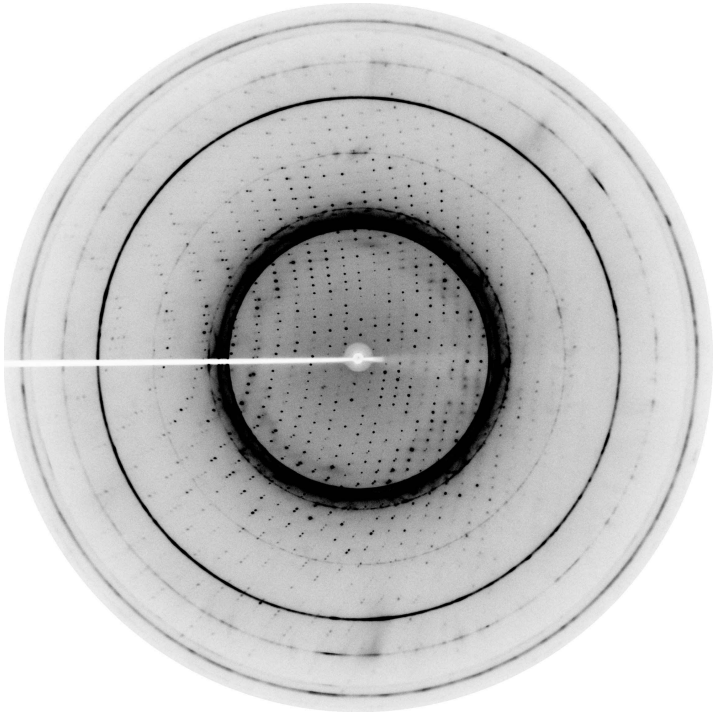

Figure 2.5: *Diffraction pattern of a BLG crystal at 1.82 Å. The experimental conditions for crystallization were 33 mg/mL BLG with 6.8 mM YCl_3 at 4 °C.*

If the amplitudes and phases of all scattered waves are known, the electron density map, which represents the atomic positions within the unit cell, can be obtained by Fourier transform[162]. Mathematically,

$$S(h,k,l) = \sum_j f_j exp[2\pi i(hx_j + ky_j + lz_j)] \tag{2.18}$$

where $S(h,k,l)$ is the structure factor, $h;k;l$ are the Miller indices, and x_j, y_j, z_j are the positions of the j^{th} atom in the unit cell and f_j is the resolution-dependent atomic scattering factor. The fundamental problem (the "phase problem") for the analysis of crystal structures is that the phase gets lost since only the intensities are measured which are obtained by taking the square of $S(h,k,l)$[145]. When the position of several atoms is known, this often allows successive identification and phase calculation[162]. One possibility is to introduce heavy atoms or atoms with anomalous scattering into the crystals and then identify their positions[145]. The shape and size of a unit cell of a three-dimensional crystal are defined by the length of three axes and the three angles between them (a, b, c and α, β, γ). If internal symmetries within a unit cell containing multiple molecules apply to the entire lattice, they constrain the parameters of the unit cell. The smallest set of structural information that is needed to reconstruct the complete crystal lattice by symmetries and lattice translations is called asymmetric unit[145].

Part II

Results & Discussion

Chapter 3

Publication A. *Nonclassical Pathways of Protein Crystallization in the Presence of Multivalent Metal Ions*

Cryst. Growth Des. ***14*** *(2014), 6357–6366.*

Andrea Sauter, Melanie Oelker, Georg Zocher, Fajun Zhang, Thilo Stehle and Frank Schreiber

Contributions	
Research design	*AS, FS, FZ*
Experiments	*AS, GZ, MO, FZ*
Data analysis and interpretation	*AS, FZ, GZ, MO*
Paper writing (structural biology part)	*GZ, MO, TS*
Paper writing	*AS, FS, FZ*

Using bovine β-lactoglobulin as a model system, we have studied the crystallization pathways in the presence of the di- and trivalent salts $ZnCl_2$ and YCl_3. Previous work has shown that protein solutions undergo a reentrant condensation (RC) phase behavior in the presence of YCl_3, i.e. a condensed phase occurs in between two boundary salt concentrations, $c^* < c^{**}$. In the presence of $ZnCl_2$, c^* also exists, but protein solutions with high salt concentrations do not become completely clear at higher protein concentrations ($> 20\,\mathrm{mg/mL}$). We thus denote this diffuse transition as *pseudo*-c^{**}. Small angle X-ray scattering measurements show that the effective interactions change from repulsion to attraction near c^*; and attractive interactions dominate around $pseudo - c^{**}$. Solutions near c^* and $pseudo - c^{**}$ provide the optimized conditions for growth of high quality protein single crystals, but with different pathways. While crystal growth near c^* follows the classical one-step nucleation pathway, crystallization around $pseudo - c^{**}$ (for zinc) or c^{**} (for yttrium) follows a nonclassical process with an intermediate phase appearing before crystallization. Furthermore, the intermediate phases strongly depend on the crystallization temperature. Samples with high salt concentrations

exhibit a typical transition temperature (T_{tr}), below which the solutions become turbid. When crystallizing near T_{tr}, the intermediate phase consists of protein clusters; below T_{tr}, the intermediate phase corresponds to macroscopic protein aggregates which can further relax into a dense liquid phase before crystallization. However, the experimental data cannot distinguish whether nucleation occurs within or outside of the intermediate phase. Possible scenarios are discussed based on the equilibrium phase behavior of colloidal systems with a short-range attraction. The crystallographic analysis of the resulting crystals shows that metal ions are an integral part of the crystal lattice. Both types of metal ions can create new protein contacts in the crystal lattice via bridging; however, yttrium creates more bridging contacts compared to zinc.

3.1 Introduction

Crystallography is the dominating method for structure determination of biological macromolecules. Its major bottleneck is the crystallization from protein solutions and growing single crystals with adequate quality. The molecular interactions in protein solutions, in particular their connection to the crystallization process, are not completely understood, and crystal growth is typically performed by trial and error experiments[43].

Recent progress in protein and colloid crystallization as well as biomineralization has shown nonclassical features in the early stage of nucleation[56,171,196,202,211]. While classical nucleation theory (CNT) predicts that molecules form nuclei in a supersaturated solution with the exact density and structure of the crystals in the final stage[7,56,172], nonclassical pathways suggest an intermediate phase (clusters or dense liquid phase) existing in between the initial solution and the final crystalline state[56,171,196,202]. The free energy landscapes of nonclassical pathways show an additional free energy minimum corresponding to the intermediate phase. Computer simulations of homogeneous crystal nucleation performed by ten Wolde and Frenkel using a colloidal model with short-range attractive interactions suggested that the presence of large density fluctuations close to the critical point of liquid–liquid phase separation (LLPS) affects the pathway for the formation of a crystal nucleus[196]. The free-energy barrier that has to be overcome is considerably lowered if the change in density occurs before the structural change[196]. While the theoretically predicted two-step nucleation near the critical point is difficult to verify experimentally, accumulated experimental observations in various systems have shown the nonclassical features of crystallization under certain conditions[30,53,54,58,202]. Even in simple hard-sphere systems, simulations and experiments suggest a two-step nucleation process with amorphous clusters as precursors[75,113,114,164,168].

A different type of cluster phase has been proposed for concentrated protein solutions that is not caused by density fluctuations: Large clusters consisting of 10^5 to 10^6 proteins have been observed in lysozyme and hemoglobin systems

outside the LLPS region[65,137,138]. An approach to find an electrostatic model for clusters with up to 10^3 molecules was pursued by Hutchens and Wang[79]. This model is, however, limited by the Debye screening length and cannot explain the experimentally observed large clusters. Nevertheless, these large protein clusters have been demonstrated to play an important role in the protein crystallization process which cannot be explained by CNT[65,97,137,138,181].

In practice, an occasionally used and successful strategy to induce protein crystallization is the addition of multivalent ions[120,132,199,218]. Divalent ions have been proven crucial for the generation of high quality crystals of insulin and other protein complexes[165,200]. We have shown previously that trivalent salts such as YCl_3 can induce a reentrant condensation (RC) phase behavior in many acidic proteins such as β-lactoglobulin (BLG)[151,183,220,222,226]. For a given protein concentration, when the salt concentration is below c^* and/or above c^{**}, protein solutions are clear. At salt concentrations between c^* and c^{**} (with $c^* < c^{**}$), protein condensation occurs including aggregation, LLPS and crystallization[224,227]. Interestingly, depending on the position in the phase diagram, classical crystallization or a temperature-dependent nonclassical crystallization process has been observed[227]. Further studies suggested that proteins form clusters in the presence of trivalent metal ions that can form either a dense liquid phase or crystals depending on the kinetics[183,223]. However, establishing a complete picture of the nucleation mechanism in the presence of the metastable intermediate phase and its role is still a challenge.

In this work, we aim to provide new insights into the nonclassical pathways for protein crystallization in the presence of multivalent metal ions. We first determine and compare the macroscopic phase behavior and crystal growth pathways for BLG with a trivalent salt, YCl_3, and a divalent salt, $ZnCl_2$. We further characterize the effective protein–protein interactions in the molecular level using small angle X-ray scattering. Finally, we explore the role of multivalent metal ions based on the high resolution crystal structure on the atomic level. We hypothesize that the difference in valency will lead to a different binding affinity of the metal ions to the protein surface[108], which may affect the resulting crystal lattice and the kinetics of crystal growth. By providing more detailed experimental observations of crystal growth under various conditions and the structural analysis of the protein crystals, we aim to address the following questions: Which crystallization pathways are followed and what is the structure and morphology of the crystals? What is the role of multivalent ions in stabilizing the crystal lattice as well as inducing the different pathways of crystallization? Are there differences between $ZnCl_2$ and YCl_3, concerning the pathway of crystallization and the ion-bridging?

3.2 Experimental Section

3.2.1 Materials and Sample Preparation

Protein and salts were purchased from Sigma-Aldrich. The guaranteed purities were 90% for BLG from bovine milk (L3908), 97% for $ZnCl_2$ (96468) and 99.99% for YCl_3 (451363). A BLG monomer consists of 162 amino acids and has a molecular mass of about 18.3 kDa. Under physiological conditions, it is found predominantly as a dimer[207]. With an isoelectric point (pI) of 5.2, it is acidic and carries a net charge of -10 e at neutral pH[44].

All samples were prepared by mixing the required amount of salt stock solution, deionized (18.2 MΩ) degassed Millipore water and protein stock solution. Stock solutions were prepared by dissolving the protein powder or salt in deionized degassed Millipore water. The concentration of BLG solutions was determined by UV absorption measurements using an extinction coefficient of $0.96\,l \cdot g^{-1} \cdot cm^{-1}$ at a wavelength of 278 nm[182]. A Seven Easy pH instrument from Mettler Toledo was used to monitor the pH of protein solutions caused by salt addition. All samples had a pH above the pI (between 6.2 and 6.9). Since most buffers can affect the phase behavior and the solubility of salts, no additional buffer was used to avoid the effect of co-salts.

The transition temperature, T_{tr}, for samples with high salt concentration was determined by turbidity measurements. Samples were prepared in a water bath at 313 K and then slowly cooled down to 278 K with a cooling rate of about 0.1 K/min. Pictures were taken every 2 K, and the temperature of the transition from clear to turbid was assigned as T_{tr}.

3.2.2 Methods

Small Angle X-ray Scattering (SAXS) and Data Analysis

SAXS measurements were performed at the ESRF, Grenoble, France at beamline ID02. The samples were measured using a flow-through capillary cell with a thickness of about 10 µm at sample-to-detector distances of 1, 2, and 5 m. The sample in the scattering volume was exchanged for every exposure. For each sample, 10 exposures of 0.1 s or 20 exposures of 0.05 s each were measured. The two-dimensional (2D) intensity pattern was corrected to an absolute scale and azimuthally averaged to obtain the intensity profiles. Afterward, the solvent background was subtracted. More detailed information on data reduction and q-resolution calibration can be found in the literature[80,129].

Small-angle X-ray scattering data can be used to obtain information on the pair interaction potential[63,74,187]. The scattering intensity, $I(q)$, for a nonspherical system, can be expressed by

$$I(q) = N(\Delta\rho)^2 V^2 P(q) \bar{S}(q). \tag{3.1}$$

where $q = \frac{4\pi}{\lambda}\sin(\theta/2)$ is the scattering vector, θ is the scattering angle, N is the number of protein molecules per unit volume in the solution, $\Delta\rho = \rho_p - \rho_s$ is the electron density difference between the solvent and the solute, and V is the volume of a single protein. $P(q)$ is the form factor of a given protein, i.e. the scattering from a single protein molecule after orientation averaging. A form factor of an oblate ellipsoid with semiaxes a and b is used to model the protein BLG. At low protein and salt concentration, the SAXS data were fitted by the crystal structure of the BLG dimer (PDB code of 1BEB[27]) using CRYSOL[186]. The structure factor $\bar{S}(q)$ is calculated using the average structure factor approximation from a monodisperse spherical system[31], with an effective sphere diameter which is calculated from a virtual sphere with the same second virial coefficient as the ellipsoid[82,223]. In the following parts and for simplicity, we use $S(q)$ to denote $\bar{S}(q)$. To describe the effective interactions in protein solution, structure factors derived from model potentials were used for data analysis. Data fitting for the determination of the interaction potential was preformed on IGOR Pro with macros developed at NIST[90]. More details of data analysis can be found in previous work[80,225].

Protein Crystallization and Crystal Structure Analysis

Crystallization was performed at 277, 293, and 312 K by the vapor diffusion hanging drop method or by batch crystallization in small glass bottles or cuvettes. The concentrations ranged from 3 to 65 mg/mL BLG and 0-50 mM $ZnCl_2$ or 0-20 mM YCl_3, respectively. The crystallization process was followed with transmission optical microscopes: a DME from Leica Microsystems Inc. in a 293 K room, an Axioskop 40 from Zeiss with an installed Nikon E995 digital camera in a 277 K room and an AxioScope.A1 from Zeiss with an included AxioCam ICc5, temperature-controlled by a water bath with a precision of 0.05 K.

X-ray analysis was performed with two typical crystals from different regions of the phase diagram (BLG-Zn-2: 10.0 mg/mL BLG and 2 mM $ZnCl_2$; BLG-Zn-20: 3.4 mg/mL BLG and 20 mM $ZnCl_2$) both grown at 277 K. Single crystals were prepared and transferred into the corresponding salt concentration including glycerol (30-34% (w/v)). The crystals were flash frozen in liquid nitrogen and stored until data collection. Data collection was performed at beamlines X06DA at the Swiss Light Source in Villigen, Switzerland, and P13 at PETRA III in Hamburg, Germany. Data processing was performed with the XDS package[87]. Structures were solved by molecular replacement as implemented in PHASER[117] using coordinates for BLG (Protein Data Bank (PDB) code 3PH5) as search template, followed by a simulated annealing refinement as implemented in PHENIX[4] to avoid model bias. Zinc ions were placed on the basis of an anomalous difference map calculated with FFT[195]. The models were refined by several cycles of model corrections with COOT[45] and reciprocal space refinement with REFMAC5[125]. Final refinement steps included TLS parametrization. Water molecules were incorporated manually with COOT. Figures were generated with PyMOL[169]. Models

were validated with MolProbity [32] and deposited to the Protein Data Bank (PDB) with accession numbers 4LZU and 4LZV. The structures of BLG crystallized in the presence of YCl_3 (PDB codes 3PH5 and 3PH6) were determined in our previous work [227].

3.3 Results

3.3.1 Phase Behavior of BLG in the Presence of $ZnCl_2$ and YCl_3

Previous work [227] showed a strongly temperature-dependent behavior for solutions of BLG with YCl_3 near c^{**}. At higher temperatures, the samples were clear

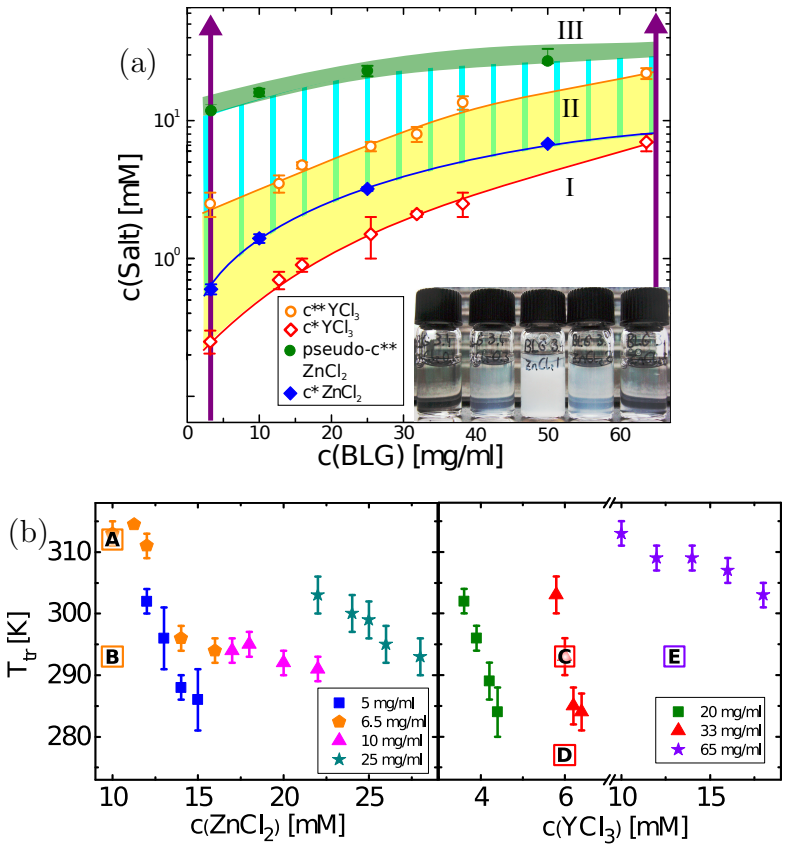

Figure 3.1: *(a) Experimental phase diagram of BLG with $ZnCl_2$ and YCl_3. In both cases, a reentrant condensation (RC) phase behavior was observed. Purple arrows mark the conditions for SAXS measurements. Inset: Photos of a series of samples with a BLG concentration of 3.4 mg/mL and increasing $ZnCl_2$ concentrations (from left to right: 0.1 mM, 0.3 mM, 1 mM, 20 mM, and 50 mM). (b) Transition temperature of BLG in the presence of $ZnCl_2$ or YCl_3. [A-E] mark samples for the study of crystal growth in the following.*

and thus in regime III. Upon cooling, they became turbid below a transition temperature T_{tr}. Similar behavior was also observed in the presence of $ZnCl_2$ for samples within the *pseudo-c*** zone. T_{tr} decreases linearly with increasing salt concentration for both salts (Figure 3.1).

3.3.2 Temperature-Dependent Crystallization of BLG in the Presence of $ZnCl_2$ and YCl_3

Crystals were observed in the second regime of Figure 3.1a with both $ZnCl_2$ and YCl_3. Using optical microscopy, we followed the crystallization procedure of BLG with different salt concentrations ($ZnCl_2$) at different temperatures. Close to (either slightly above or below) c^*, our observations suggest that the crystallization follows the classical one-step pathway; i.e., crystals nucleate directly from a supersaturated solution and grow larger without any visible cluster or LLPS phase, similar to experiments with YCl_3 [227]. Under our experimental conditions, this crystallization pathway did not depend on the crystallization temperature.

Close to (*pseudo*)-*c*** (i.e., the second boundary for both $ZnCl_2$ and YCl_3), different growth processes depending on temperature were observed. A typical example is shown in Figure 3.2 for a sample close to its transition temperature (6.5 mg/mL BLG and 10 mM $ZnCl_2$ at 312 K $\sim T_{tr}$, [a] in Figure 3.1b). No macroscopic LLPS was visible, but massive amorphous clusters appeared before crystallization. The amorphous clusters are metastable with respect to the crystals: they dissolved and disappeared at the end of crystallization.

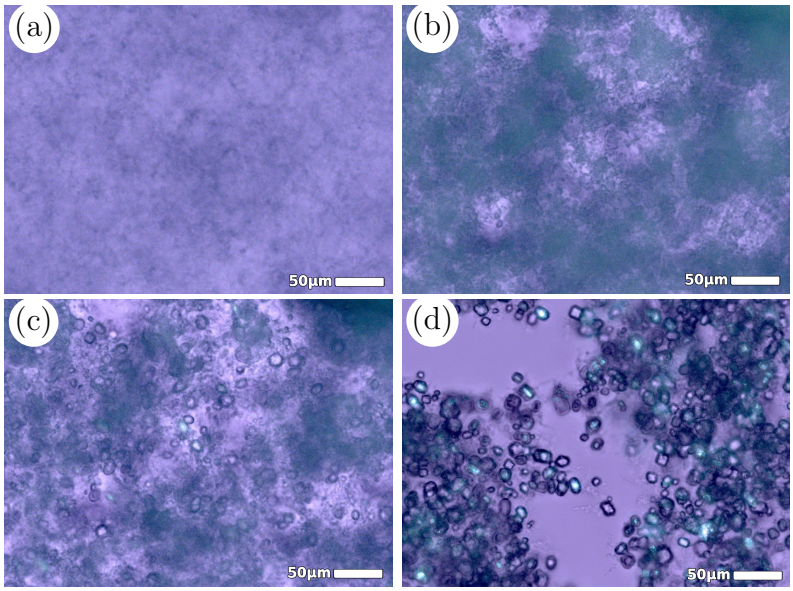

Figure 3.2: *Typical crystallization pathway in the pseudo-c** zone near T_{tr}. Crystals grown at 312 K: 6.5 mg/mL BLG with 10 mM $ZnCl_2$. (a) Directly after preparation at 313 K. (b-d) After 7, 10 and 24 h, respectively.*

For the same sample at 293 K, well below T_{tr} (Figure 3.3, label [b] in Figure 3.1b), a different growth process was observed. With time, a network of clusters forms and further relaxes into a liquid-like structure, similar to the viscoelastic behavior described in literature for polymers and for lysozyme[93,191,192]. Based on the fact that these areas are growing and merging, we conclude that they are liquid, not gel. By focusing on different depths of the sample solution, one could see that this structure forms predominantly at the lower glass interface. However, this is mainly due to the higher density of the liquid-like network structure instead of preferential wetting on the glass slide as demonstrated by the hanging drop experiments (see Supporting Information, Figure A.1). Furthermore, crystal growth mainly starts from the interface between the dense and the dilute liquid phases. During crystal growth, the dense liquid phase is consumed.

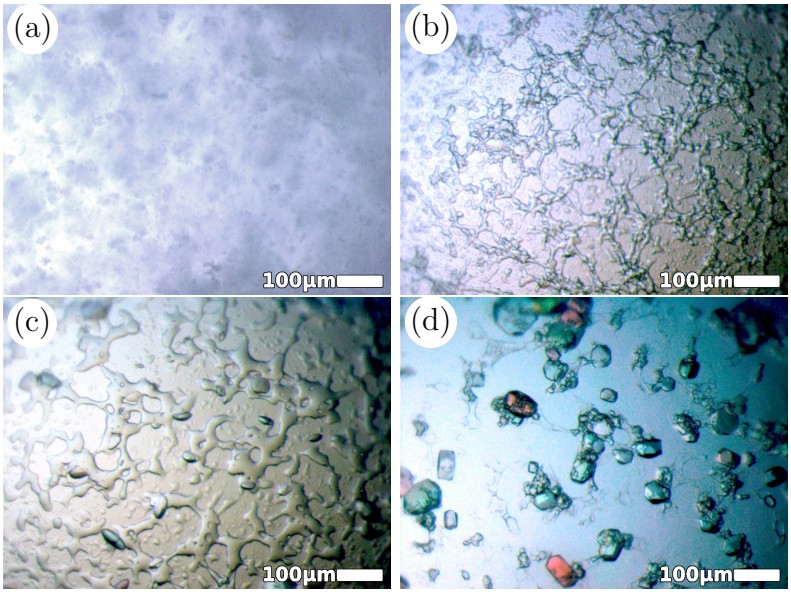

Figure 3.3: *Typical crystallization pathway in the pseudo-c** zone below T_{tr}. Crystals grown at 293 K at the liquid-glass interface: 6.5 mg/mL BLG with 10 mM $ZnCl_2$. (a) Turbid sample directly after preparation. (b) Network-like dense liquid structure after 2 h. (c) After 7 h, the dense regions grew and crystals were visible. (d) After 24 h, crystal growth had consumed all dense liquid material.*

Previously, we were able to show the strong dependence of protein crystallization on salt concentration (YCl_3) and temperature[227]. Here, we give an example of the same sample near c^{**} that follows different pathways by varying temperatures below T_{tr}, similar to the experiments with $ZnCl_2$ described above. In Figure 3.4, the crystal growth process of a sample with 33 mg/mL BLG and 6.0 mM YCl_3 at 293 K (labeled [c] in Figure 3.1b) is shown. This temperature is close to T_{tr} and the sample was slightly turbid at first, but no macroscopic LLPS was visible (Figure 3.4a). Crystallization occurred after about 3 days (Figure 3.4b). With crystal growth, the turbid solution became clear and only crystals were visible in

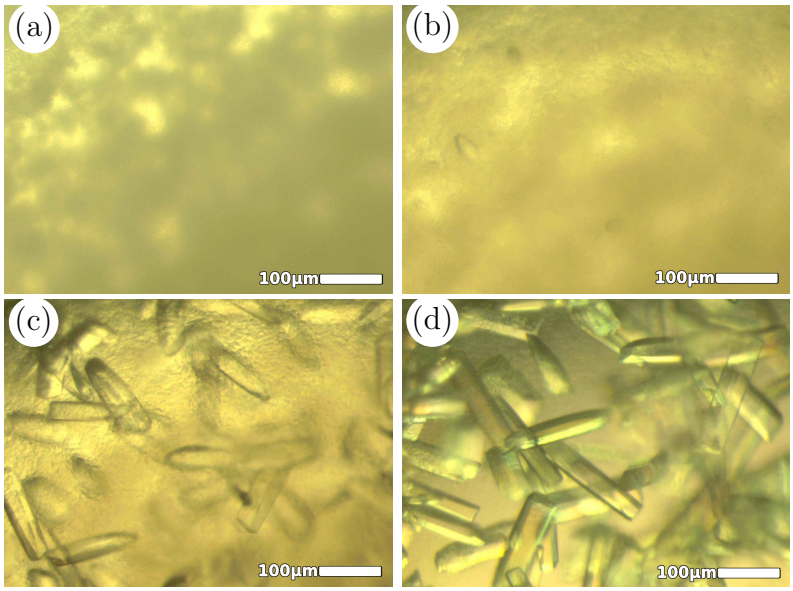

Figure 3.4: *Pictures originate from samples prepared by the hanging drop method without prior centrifuging. Crystal growth at 293 K with a protein concentration of 33 mg/mL and 6.0 mM YCl_3, close to the transition temperature. No liquid phase was visible but turbidity caused by large clusters. (a) After 3.5 h. (b) After 75 h. (c) After 101 h. (d) After 165 h.*

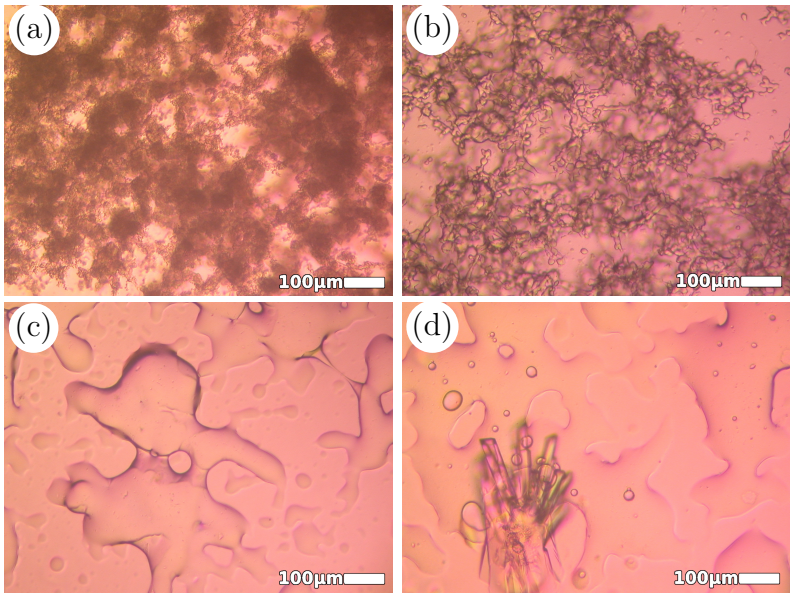

Figure 3.5: *Formation of droplets at the sample-glass interface in a cuvette and crystal growth at 277 K. The sample has a concentration of 33 mg/mL BLG and 6.0 mM YCl_3. (a) Directly after preparation: turbid sample. (b) After 2 h: fine liquid network. (c) After 23 h: large dense liquid regions. (d) After 54 h: crystals.*

the end. No crystal growth was observed at temperatures above T_{tr}. Well below T_{tr}, another process was observed (Figure 3.5). When quenched to 277 K (sample [d]), the solution becomes turbid quickly and massive amorphous clusters are clearly visible (Figure 3.5a). With time, these clusters formed a liquid network (Figure 3.5b and c). Crystal growth was observed mainly at the interfaces and the dense liquid-like network is consumed in the end. A similar crystallization pathway was also observed for samples with higher protein concentrations. An example is shown in Figure 3.1b. The sample contained 65 mg/mL BLG and 13 mM YCl_3, the transition temperature is about $T_{tr} = 310$ K. When crystallized at 293 K (sample [e]), a liquid-like network as an intermediate step was observed (shown in the Supporting Information, Figure A.2). This behavior is very similar to that observed in Figure 3.3.

The crystallization procedures at different temperatures are inconsistent with the classical nucleation theory. The appearance of an intermediate phase (macroscopic LLPS or microscopic clusters) before crystallization indicates that an additional free energy minimum exists in between the supersaturated solution and the final crystalline state. Since the resolution of the optical microscope used in this work cannot distinguish whether the nucleation starts directly from the dense liquid phase or the cluster phase (or rather takes place as heterogeneous nucleation at the droplet surface), we simply denote the observed crystallization processes as different *nonclassical* pathways. The interesting observation here is that for the same sample below the transition temperature, depending on the depth of quench, the intermediate phase of the crystallization can be microscopic clusters or macroscopic LLPS. The physical mechanism of this observation will be discussed after the structural analysis of the resulting crystals.

3.3.3 Effective Protein–Protein Interactions Revealed by SAXS

SAXS was used to characterize the effective interactions in protein solutions under conditions similar to the crystallization conditions. We first determined the form factors of proteins in different regimes using samples with low protein concentration, i.e., 3.3 mg/mL (Figure 3.6a). At very low salt concentrations, the SAXS data can be well fitted using the crystal structure of the BLG dimer (PDB code of 1BEB) using Crysol[186], which can be simplified by an ellipsoid form factor with axis a=38 Å and b=19 Å. The respective fitting parameters are shown in the Supporting Information, Table A.1. It was found that the values of a and b were nearly constant and $a > b$ in the first regime (i.e. for 0.1 and 0.3 mM $ZnCl_2$). With 20 and 50 mM $ZnCl_2$, the values of b became larger and $a < b$, indicating the formation of oblate-like clusters in the third regime. The minimum in scattering intensity at a q of $\approx 0.18 - 0.25$ Å^{-1} in Figure 3.6a and b, is dominated by the form factor. This is in contrast to the data from BLG solutions in the presence of YCl_3[223], where a peak was observed at 0.22 Å^{-1} that was assigned to the

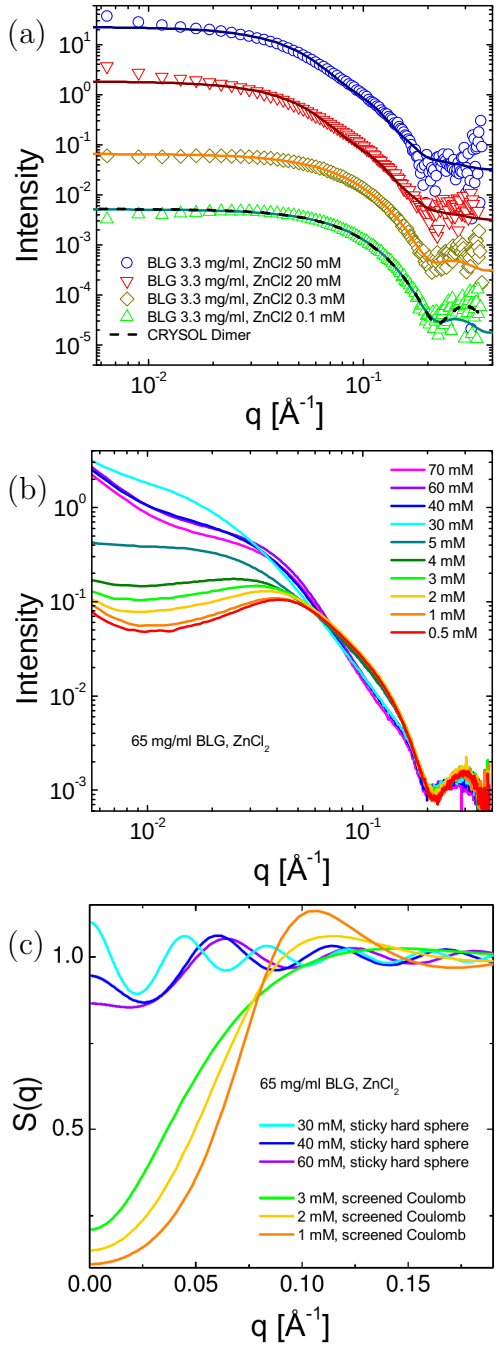

Figure 3.6: *SAXS measurements at 3.3 mg/mL and 65 mg/mL BLG with various $ZnCl_2$ concentrations. For better visibility, the curves were shifted in intensity and only every third data point is shown. (a) Fitted form factors for 3.3 mg/mL BLG (continuous lines: ellipsoid form factor, dashed line: form factor from crystal structure). (b) SAXS measurements at 65 mg/mL BLG with $ZnCl_2$ concentrations in the first and third regime. (c) Calculated structure factor of 65 mg/mL BLG with part of the $ZnCl_2$ concentrations from b.*

monomer-monomer correlation within the clusters. The reason for this difference may be the stronger bridging effect of yttrium in comparison to zinc as explained in the following section.

Samples in the condensed regime were also measured, but due to the precipitation of proteins, these results were not used for further analysis.

SAXS data for samples with higher protein concentrations were analyzed for the effective interactions. Figure 3.6b and c present the SAXS data and calculated structure factor for samples with 65 mg/mL BLG and various salt concentrations. The corresponding c^* and $pseudo - c^{**}$ are around 7 mM and 30 mM, respectively. At low salt concentration, i.e. well below c^*, the SAXS data can be well described by a combination of an ellipsoid form factor and a screened Coulomb structure factor[73,77], indicating repulsive interactions between protein molecules. The strength of repulsion decreases with increasing salt concentration and the fitted surface charge continuously decreases upon approaching c^* (Supporting Information, Table A.2). In the third regime, an ellipsoid form factor with a sticky hard sphere structure factor was used[15], which represents a short-ranged attractive interaction potential. The obtained volume fractions are close to the calculated value and the interaction strength decreases if the salt concentration is further increased from $pseudo - c^{**}$.

In summary, our SAXS measurements suggest that the effective interactions are repulsive and dominated by the surface charges of proteins at low salt concentrations ($ZnCl_2$). With increasing salt concentration, the repulsive potential is reduced and the potential becomes attractive approaching c^*. In the third regime, the effective interactions are still attractive, but decrease in strength with increasing salt concentration in good agreement with the visual observation that the solutions become less turbid. Because the attractions in the condensed regime are too strong for the growth of high quality crystals, the SAXS measurements indicate that the optimized conditions of crystallization would be near the two boundaries. These results are consistent with previous studies of BLG solutions in the presence of YCl_3[223].

3.3.4 Crystal Structures and the Role of Multivalent Metal Ions

We determined the structures of two crystals (BLG-Zn-20, BLG-Zn-2) to resolutions of 2.44 Å and 2.40 Å respectively, including the zinc positions that could be determined unequivocally. Both crystals belong to space group $P3_2\,2\,1$ and contain one monomer in the asymmetric unit. The 2-fold axis of the biological dimer is aligned with the crystallographic axis. Although the same space group has been reported for BLG crystals earlier[146], this crystal form is new and differs in the unit cell dimensions. The crystal packing of BLG with zinc ions is shown in Figure 3.7a. Data collection and refinement statistics of both structures are given in the SI in Tables A.3 and A.4.

The positions of zinc ions were determined by an anomalous density map using the Bijvoet differences as Fourier coefficients and relative occupancies were determined during the refinement process. In both cases, one crystal contact is mediated by a zinc ion. Furthermore, a different number of additional zinc ion binding sites is not contributing to the crystal lattice formation.

The analysis of the cation binding sites in BLG-Zn-2 shows that the crystal contact is mediated by a zinc ion (zinc 1 in Figure 3.7), which is coordinated by residues D53 and E74 from one protomer and E158 and H161 from a symmetry related protomer (Figure 3.7b). An additional zinc binding site is found at H146, without mediating a crystal contact. In contrast to the fully occupied zinc 1 binding site, this site is only partially occupied (30%) at a concentration of 2 mM $ZnCl_2$. An image is found in the SI in Figure A.4.

The zinc ions in BLG-Zn-20 mediating a crystal contact are coordinated by the same residues found in BLG-Zn-2 with the exception of E158 (Figure 3.7c). This residue is located at a different position. The zinc binding site at H146 (shown in Figure 3.7d) is found in BLG-Zn-20 as well, but the occupancy is higher (75%) compared to BLG-Zn-2 which is in good agreement with the $ZnCl_2$ concentration. Furthermore, 3.3 Å away from the carboxy group of E158, a peak in the anomalous map indicates the presence of an additional zinc site. Modelling of a zinc ion with low occupancy resulted in long binding distances to E158 (shown in Figure 3.7e) which indicates a very loose metal binding site and was therefore not included in the final structure coordinates. A second additional binding site can be found at E51 with an occupancy of ca. 70% (shown in Figure 3.7c).

A packing analysis of both structures was performed by PISA [91]. In addition to the dimeric interface, which buries solvent accessible areas of 514 $Å^2$ and 590 $Å^2$ for BLG-Zn-2 and BLG-Zn-20, respectively, additional proteinaceous contacts build up the crystal lattice ranging from 35 $Å^2$ to 278 $Å^2$. In case of 2 mM $ZnCl_2$, the crystal contact mediated by zinc 1 has an interface of 73 $Å^2$ and in case of 20 mM $ZnCl_2$, the area has a size of 77 $Å^2$. For the additional zinc binding sites, the interface areas are 42 $Å^2$ for zinc 2 in the structure with 2 mM $ZnCl_2$ and 40 $Å^2$ and in case of 20 mM $ZnCl_2$ 39 $Å^2$ for zinc 2 and 3, respectively.

In previous study, we determined structures of BLG crystallized in the presence of different concentrations of YCl_3 on the basis of its phase diagram (PDB codes 3PH5 and 3PH6, Ref. [227]). We reported four yttrium ions in the BLG-Y structures, all of them were coordinated by acidic side chains and contribute to the lattice formation. One yttrium site (yttrium 4) showed substantially lower occupancy when lower concentrations of YCl_3 were applied in the crystallization process. BLG-Zn crystals share some similarities with BLG-Y, but also strong differences. As in BLG-Y, the bridging metal ions in BLG-Zn are coordinated mainly by acidic side chains, which is obvious, as acidic side chains compensate for the positive charge of the metal ions. In BLG-Zn, zinc ions are moreover coordinated by histidine residues, e.g. H161 at zinc 1. Coordination of zinc by histidine side chains is frequently observed in nature, e.g. in the family of zinc finger proteins (Ref. [170]). In contrast to BLG-Y, where four sites built up a three-dimensional

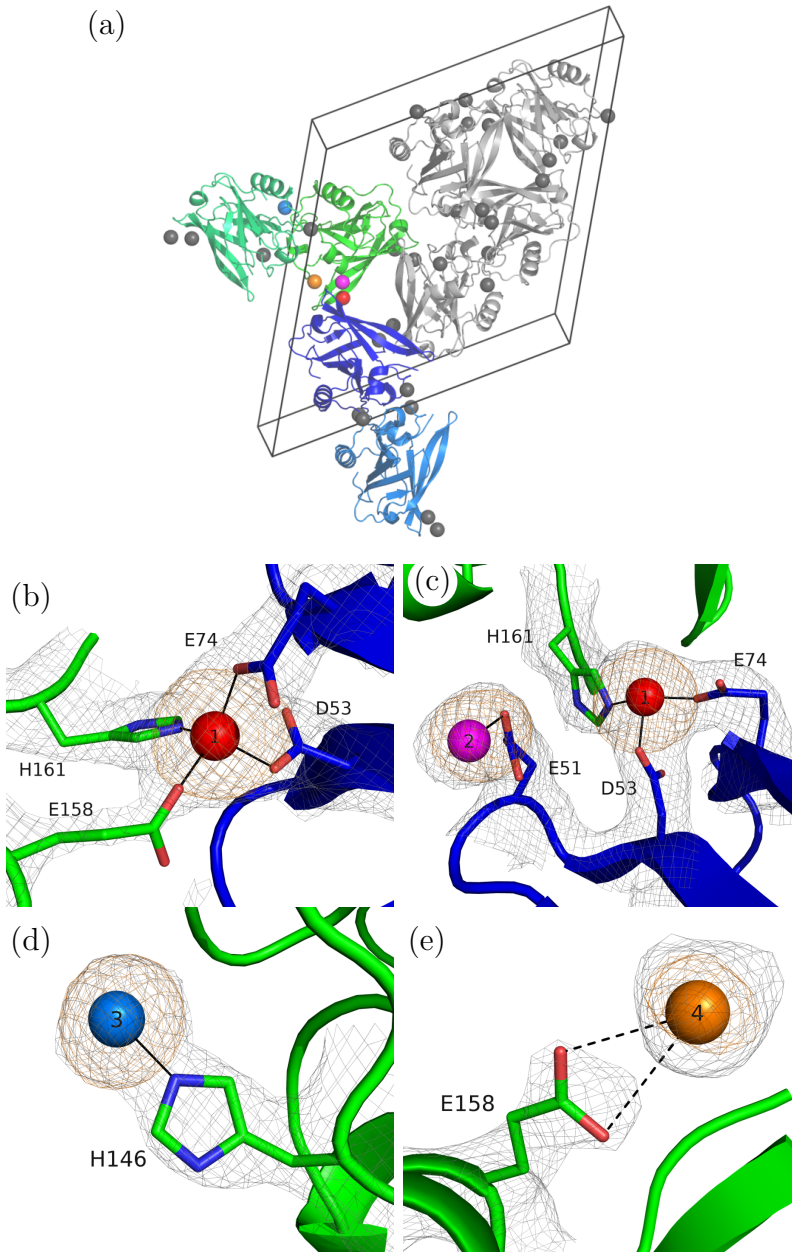

Figure 3.7: *(a) Crystal packing of BLG crystallized with 2 mM and 20 mM $ZnCl_2$. Two representative dimers are colored blue and green. Asymmetric units are represented by different color shades. Other monomers in the unit cell are represented in gray. Zinc ions are presented as gray spheres. Representative zinc ions are colored red (zinc ion mediating the crystal contact of the two representative dimers) and blue. Additional zinc sites identified in BLG-Zn-20 exclusively are colored magenta and orange. (b-e) Cation binding sites in BLG-Zn-2 (b) and BLG-Zn-20 (c-e). Density maps were calculated with FFT[195]. Sigma levels of the maps are $\sigma = 1.0$ for $2F_{obs}$-F_{calc} maps (gray) and $\sigma = 4.0$ for anomalous maps (orange). Zinc 4 is not included into the coordinates due to its long binding distances.*

Table 3.1: *Comparison of BLG-Y and BLG-Zn Structure*

	space group	binding sites	bridging sites
BLG-Y	$P2_1\,2_1\,2_1$	4	4
BLG-Zn	$P3_2\,2\,1$	4	1

crystal lattice, we identified a single zinc ion contributing to the lattice formation (Figure 3.7b, Table 3.1). Another zinc site was identified, coordinated solely by H146 (Figure 3.7d), and it did not contribute to protomer-to-protomer interactions. Interestingly, this site revealed a concentration-dependent effect similar to yttrium 4 in BLG-Y. Moreover, at higher zinc concentrations an additional metal ion was identified. A highly speculative explanation for the reduced amounts of bridging metal ions might be found in the decreased Lewis acid character of zinc ions compared to yttrium ions, which might hamper the intermolecular bridging effect. This is in good agreement with the observation that main crystal contacts in BLG-Zn derive from proteinaceous interactions. In summary, crystallographic analysis indicates that the bridging effect of high valency metal ions promotes new protein contacts and stabilizes crystal lattice with both metal salts.

3.4 Discussion

We have determined the phase behavior of the globular protein BLG in the presence of the divalent salt $ZnCl_2$. The phase behavior is similar to that observed for BLG solutions in the presence of trivalent salts, such as YCl_3. However, the RC behavior in the presence of $ZnCl_2$ is not generally found for other acidic proteins examined, such as BSA and HSA. Despite this specificity, the solutions near the second phase boundary show similar temperature-dependent phase transitions. Moreover, while crystallization follows the classical one-step pathway near c^* with both salts, the crystal growth near (*pseudo*)-c^{**} follows nonclassical pathways depending on the crystallization temperature. These findings and their possible interpretation are summarized in Figure 3.8. In the following, we discuss our experimental observations based on the established phase behavior in model colloidal systems.

Colloidal systems with short-range attraction exhibit a phase diagram (T, c) with a metastable LLPS[7]. Although such a phase diagram is to be determined for the investigated systems, the observed temperature-dependent intermediate phases can still be explained using Figure 3.8a. At salt concentrations close to c^* (top), the crystallization follows a classical one-step process and does not depend strongly on temperature. The hidden LLPS, if it exists, is far below the gas-crystal (G-C) phase boundary (outside the observed temperature range) under these conditions. Lowering the temperature for crystallization within the G-C region does not change the nucleation behavior. This pathway is labeled as "1". At high salt concentrations, close to (*pseudo*)-c^{**} (bottom), the existence of an intermediate

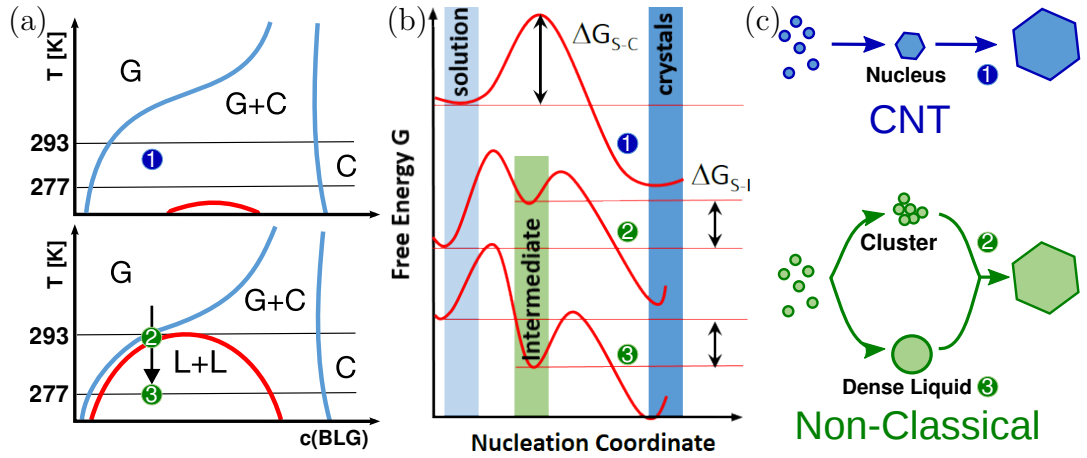

Figure 3.8: *(a) Possible location of investigated samples in a typical phase diagram (T, c plane) for colloidal systems with a short range attraction. Samples near c* may have a metastable LLPS region far below the G-C boundary, whereas samples near (pseudo)-c** have a LLPS region very close to the G-C boundary (in both BLG-Y and BLG-Zn). Adapted and modified from Ref.*[7]*. (b) Free energy landscape of the different crystallization pathways. The additional minimum corresponds to the formation of intermediate phases during the nonclassical crystallization. Adapted and modified from Ref.*[202]*. (c) Summary of the observed crystallization pathways. "Path-1" represents the classical one-step pathway which is followed by samples near c*. "Path-2" and "Path-3" represent nonclassical pathways with an intermediate phase of protein clusters and macroscopic LLPS, respectively, that are followed close to (pseudo)-c**. Note that these nonclassical pathways are not necessarily the same as the "two-step" mechanism proposed in theory*[196]*. Details are discussed in the text.*

phase suggests that the liquid-liquid (L+L) coexistence region is very close to the G-C phase boundary. When quenching the solution only slightly below the transition temperature, the sample is in the state labeled as "2" in the sketched phase diagram. This position is slightly below the G-C boundary, but importantly, it is also very close to the binodal of LLPS. Under these conditions, density fluctuations lead to the formation of a microscopic protein cluster phase before crystal nucleation occurs. This scenario ("Path-2") is shown in Figures 3.2 and 3.4. At higher temperature, the gas phase region is reached and no crystallization can be found anymore. If the sample is quenched deeper, i.e. to location "3" which is well below the LLPS boundary, then macroscopic metastable LLPS occurs as the intermediate phase before nucleation. This scenario ("Path-3") has been shown in Figure 3.3 and 3.5.

The corresponding free energy landscapes of these three pathways are described in Figure 3.8b. In the classical pathway associated with "Path-1", there is one free energy maximum, an energy barrier that has to be overcome in order to nucleate. The free energy landscapes of the nonclassical pathways show an additional min-

imum corresponding to the intermediate phase. If the free energy of the intermediate phase is higher than the free energy of the initial solution, the intermediate phase is unstable and consists of mesoscopic clusters. This could be the case for samples close to the LLPS boundary like "Path-2". If the free energy of the intermediate phase is lower than that of the initial solution, the metastable phase is a macroscopic dense liquid phase, e.g. "Path-3". This has been summarized in Ref.[202].

In Figure 3.8c, the observed crystallization pathways are illustrated. For samples prepared close to c^*, nucleation occurs directly in the supersaturated solution with no intermediate structure or phase transition. This classical pathway is referred to as "Path-1". Near (*pseudo*)-c^{**}, two different nonclassical pathways are observed depending on the crystallization temperature. "Path-2" describes the crystallization of samples located near the LLPS binodal in Figure 3.8a where density fluctuations lead to the microscopic clusters before nucleation. "Path-3" describes that protein solutions first undergo a macroscopic LLPS as it locates deep inside of the LLPS region before crystallization starts. However, the existence of an intermediate phase does not mean that the nucleation occurs within protein clusters or the dense liquid phase. Our observations as well as other experimental studies[30] suggest that crystallization takes place rather at the interface of dense and dilute regions than within the dense phase. These nonclassical pathways observed experimentally are not the same as the two-step process first predicted by ten Wolde and Frenkel[196]. The theoretically predicated "two-step" mechanism describes one nucleation event in which the two order parameters (density and structure) develop in two different time scales, i.e. the supersaturated solutes first undergo a density development and form clusters or macroscopic coexistent liquid phases. In the second, subsequent step, the clusters or the dense liquid phase with already higher density undergoes a structure change and complete the nucleation event. In practice, however, another nonequilibrium phase behavior in the dense phase, i.e. gelation, may prevent the second step of structure formation as any conformation change in a gel state will be extremely slow.

On the other hand, when looking at the phase diagram in Figure 3.8a (bottom), because the G-C and L-L phase boundaries are very close, samples located in this region face two potential phase transitions, i.e. both crystallization and LLPS. Kinetically, LLPS occurs faster than crystal nucleation as the corresponding energy barrier is lower. Energetically, LLPS is metastable with respect to the crystalline state. From this point of view, one cannot completely rule out the possibility of two parallel classical phase transitions. Further studies should focus on real-time characterization of the role of the intermediate phase on the nucleation process, the early nucleation, and growth kinetics of such a nonclassical crystallization process.

3.5 Conclusion

We have studied the influence of the di- and trivalent electrolytes $ZnCl_2$ and YCl_3 on protein (BLG) crystallization on levels from macroscopic phase behavior to the effective intermolecular interactions and further down to the atomic crystal structures.

A reentrant condensation phase behavior has been observed in protein solutions with both salts, i.e., a condensed regime exists between c^* and (*pseudo*)-c^{**}. The reentrant effect is weaker with $ZnCl_2$, which might be mainly due to its lower valency compared to YCl_3 and other trivalent salts. SAXS measurements indicate that both salts can tune the effective protein–protein interaction from repulsive to attractive by crossing c^*. Above (*pseudo*)-c^{**}, the interactions are still attractive but decrease with further increasing salt concentration. The effective attractions near c^* and (*pseudo*)-c^{**} provide the optimum conditions for protein crystallization.

Crystal growth in these areas followed by optical microscopy show different scenarios: while crystal growth near c^* follows a classical one-step pathway, two different nonclassical pathways are observed close to (*pseudo*)-c^{**} depending on the crystallization temperature. If the crystallization temperature is close to the transition temperature, the intermediate phase is a microscopic cluster phase. If the sample is quenched well below the transition temperature, the intermediate phase corresponds to a macroscopic LLPS. These temperature-dependent nonclassical pathways are discussed based on the phase behavior of colloidal systems with a short ranged attraction where a metastable LLPS region exists below the G-C coexistence line. Furthermore, crystallographic analysis of the resulting high quality protein single crystals demonstrates that the multivalent metal ions provide new contacts within the lattice via bridging.

We gratefully acknowledge the financial support from DFG and the allocation of beamtime as well as the support of the beamline scientists at SLS, ID02 (ESRF) and PETRA III. A.S. acknowledges a fellowship of the Landesgraduiertenförderung. The authors thank Hans-Joachim Schöpe for valuable discussions and comments.

Chapter 4

Publication B. *Real-Time Observation of Nonclassical Protein Crystallization Kinetics*

J. Am. Chem. Soc. ***137*** *(2015), 1485–1491.*

Andrea Sauter, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Robert M. J. Jacobs and Frank Schreiber

Contributions	
Research design	*AS, FS, FZ*
Experiments	*AS, FZ*
Technical Assistance	*GL*
Rate equation model	*FRR*
Data analysis and Interpretation	*AS, FRR, FS, FZ, RMJJ*
Paper writing	*AS, FRR, FS, FZ*

We present a real-time study of protein crystallization of bovine β-lactoglobulin in the presence of $CdCl_2$ using small angle X-ray scattering and optical microscopy. From observing the crystallization kinetics, we propose the following multistep crystallization mechanism that is consistent with our data. In the first step, an intermediate phase is formed, followed by the nucleation of crystals within the intermediate phase. During this period, the number of crystals increases with time, but the crystal growth is slowed down by the surrounding dense intermediate phase due to the low mobility. In the next step, the intermediate phase is consumed by nucleation and slow growth, and the crystals are exposed to the dilute phase. In this stage, the number of crystals becomes nearly constant, whereas the crystals grow rapidly due to access to the free protein molecules in the dilute phase. This real-time study not only provides evidence for a two-step nucleation process for protein crystallization but also elucidates the role and the structural signature of the metastable intermediate phase in this process.

4.1 Introduction

Protein crystallization is the bottleneck of many projects in structural biology, since, in general, suitable crystallization conditions cannot be reliably predicted [127]. In classical nucleation theory, the solute molecules reversibly aggregate in the supersaturated solution and form nuclei with the density and structure of the crystals in the final stage. While classical nucleation theory has been successfully used in many systems, including protein crystallization under certain conditions [180], various studies in the crystallization of proteins, colloids and clathrate hydrates as well as biomineralization have shown features beyond the classical view in the early stage of nucleation [34,57,58,83,171,211].

In contrast to atomic systems, proteins and colloids often have an attractive interaction short-ranged compared to the size of the particles, as demonstrated by experiments [30,64,65,100,137,138], theory [79,198], and simulations [164,196]. Such short-ranged attraction leads to a liquid–liquid phase separation (LLPS) which is metastable with respect to the crystalline phase [7,70,142,196,202]. For these systems, a two-step (or multistep) mechanism of crystal nucleation has been discussed, i.e. an intermediate state, either cluster, aggregates or a dense liquid phase, exists between the initial solution and the final crystalline state and may serve as the nucleation precursor. We note that throughout this paper, we use the term 'aggregates' instead of 'clusters' to avoid misinterpretation. An important theoretical prediction is that nucleation is favored in the dense liquid intermediate phase compared to the dilute phase, since the surface energy of the dense phase is closer to the final crystalline state, and the free energy barrier for nucleation is lower [171,196,202]. However, experimental observations in several protein systems suggest that crystals nucleate mainly from the dilute phase or at the interface of the dense liquid droplets and grow outside of it into the dilute phase [30,100,148]. Despite the existing various theoretical concepts, clear experimental evidence and suitable methods to distinguish the early stage of nucleation are rather rare [161,168,181,228]. Early studies of protein crystal growth in real space using atomic force microscopy have revealed many important features of the metastable protein clusters and their role in the nucleation process [94,111,119,181,201]. However, the quantitative real-time *in situ* characterization of the transition kinetics and microscopic structure from the metastable intermediate phase to the stable crystalline phase is still a challenge.

Our previous studies have shown that trivalent salts such as YCl_3 lead to a reentrant condensation and metastable liquid–liquid phase separation phase behavior in many solutions of negatively charged proteins [220,224,226,227]. Importantly, crystal growth depends on the exact location in the phase diagram: Depending on external control parameters such as salt concentration and temperature, classical crystallization or a temperature-dependent nonclassical process were observed, but could not be analyzed in further detail regarding kinetics and nucleation mechanism. The optimum crystallization conditions were found close to the reentrant condensation phase boundaries [227].

In this work, we investigate the crystallization *kinetics* of the globular pro-

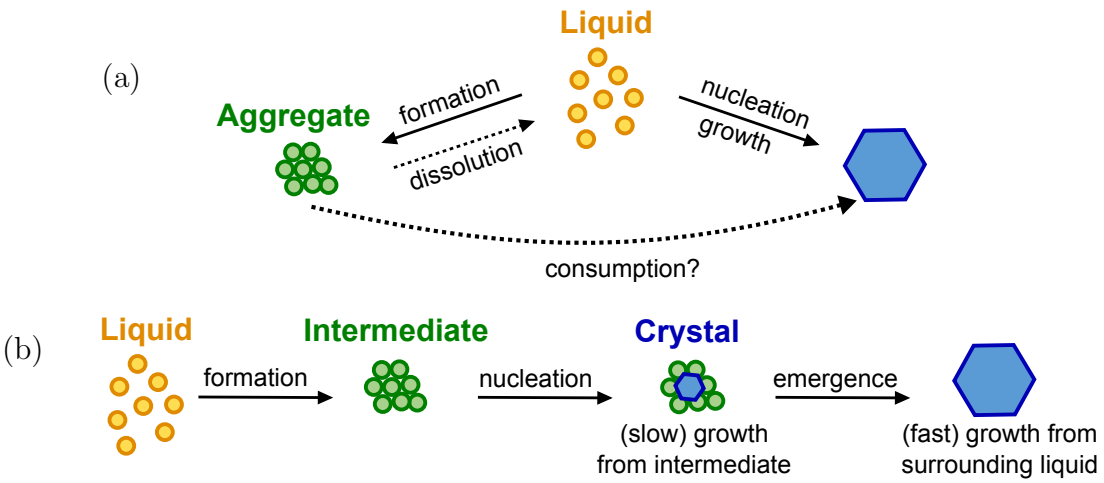

Figure 4.1: *Schematic of all processes in two possible pathways of crystallization. (a) Nucleation directly from the liquid, parallel formation of "intermediates" that are redissolved or consumed by the growing crystals. (b) Intermediates are formed first and act as precursor for crystal nucleation.*

tein β-lactoglobulin (BLG) in the presence of the divalent salt $CdCl_2$ by real-time SAXS and optical microscopy. BLG has turned out to be a good model system for crystallization by tuning the protein interactions utilizing reentrant condensation with trivalent salts. Here, we focus on the effect of the formation of aggregates on the crystallization process and which crystallization pathways are followed. In a distinct phase diagram region, protein aggregates form before crystallization. Given the dimensions of proteins, a direct optical visualization of the nucleation process in protein solutions is not possible. Here, we employ the method of real-time SAXS, monitoring the nucleation process by means of structural information on the solutes such as protein monomers, crystals, and intermediates. These experiments raise interesting questions regarding protein nucleation. A model is proposed to explain the main features of the experiments. Different scenarios of crystal formation can be compared to the kinetic data. While the classical one-step nucleation describes the direct nucleation and growth from a homogeneous solution, in particular two nonconventional pathways are discussed (see Figure 4.1): first, a *parallel* process of one-step crystal nucleation accompanied by an "intermediate" that forms and dissolves in the solution depending on concentration (Figure 4.1a). In this pathway, crystal growth takes place via the liquid or possibly by incorporation of aggregates, which, however, are not essential for crystal nucleation. Second, a *two-step* process of crystal nucleation within an intermediate that forms first from solution and is later transformed to crystals (Figure 4.1b).

4.2 Experimental

Lyophilized protein powder of β-lactoglobulin (BLG) from bovine milk and the divalent salt cadmium chloride ($CdCl_2$) were purchased from Sigma-Aldrich (product no. L3908 and 202908) and used as received.

Stock solutions were prepared by dissolving the protein powder or salt in deionized (18.2 MΩ) and degassed Millipore water. All samples were prepared by mixing the needed amount of protein, salt stock solution, and Millipore water. The protein concentration was determined by UV absorption using an extinction coefficient of 0.96 $mL \cdot mg^{-1} cm^{-1}$ at a wavelength of 278 nm[182]. To avoid the effect of buffer on the phase behavior, no additional buffer was used for sample preparation. The pH of the solutions was monitored using a Seven Easy pH instrument from Mettler Toledo and was in between 6.0 and 7.0. Compared with the $pI = 5.2$ of BLG[44], the pH effect of added salt could not lead to the charge inversion of proteins[151,226]. All experiments were performed at 293 K.

Circular dichroism (J-720 spectrophotometer from Jasco Inc.) and Fourier transform infrared spectroscopy (FTIR) (IFS 48 from Bruker) were applied to monitor the stability of the protein secondary structure under the experimental conditions.

Crystal growth was followed by optical microscopy (DME from Leica Microsystem GmbH or AxioScope A1 from Zeiss). Before crystallization, protein stock solutions were filtered (100 μm). The crystallization process was performed using either a vapor diffusion hanging drop method or a batch method between two narrow glass slides (thickness approximately 300 μm), sealed by silicone. Both methods gave similar results.

SAXS measurements were carried out at the ESRF, Grenoble, France at beamline ID02. With a sample-to-detector distance of 2 m and an energy of 16047 eV, the accessed q-range was 0.06-4.3 nm^{-1}. Details of the beamline, data collection, and calibration can be found in Ref. [129]. For time-resolved SAXS measurements, freshly prepared sample solutions in a quartz capillary were quickly transferred to the sample station and the measurement started about 2-3 min after mixing. SAXS data were collected every 2-5 min depending on the samples. The exposure time was 0.05 s. To check the consistency of the measurements, the sample was shifted vertically to different positions. This cycle was repeated every 2-5 min. Only the data from one position as a function of time was used for the data analysis shown below, but data from all positions are consistent with our analysis.

4.3 Results

4.3.1 Experimental Phase Diagram and Protein Stability

To provide a basis for the kinetic studies, we first present the experimental phase diagram of BLG in the presence of $CdCl_2$ (Figure 4.2). Upon adding $CdCl_2$,

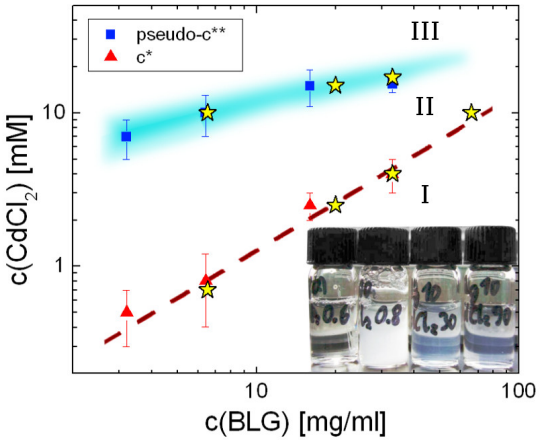

Figure 4.2: *Experimental phase diagram of BLG and $CdCl_2$ at room temperature. Yellow stars mark the conditions used for the SAXS experiments. Inset: Samples with a fixed BLG concentration (10 mg/mL) and increasing $CdCl_2$ concentrations (from left to right: 0.6, 0.8, 30, and 90 mM).*

BLG solutions show a sharp transition c^* from a clear solution at low $CdCl_2$ concentrations (regime I) to a turbid one with massive precipitates at higher $CdCl_2$ concentrations (regime II). At further increased salt concentrations, the precipitates are gradually dissolved and the solution becomes less turbid, but not completely clear even for a very high salt concentration. This is in contrast to systems with trivalent salts where we observed a second border c^{**} above which samples become completely clear again (reentrant effect)[220,226]. We define this partial transition zone to clear solutions as *pseudo-c***. The region above *pseudo-c*** is denoted as regime III. Zeta potential measurements show that a $CdCl_2$ concentration-dependent charge inversion of the protein takes place (S.I., Figure B.1b). In Figure 4.2, stars mark the conditions used for the real-time SAXS measurements. Under the current experimental conditions, no liquid–liquid phase separation was observed in BLG solutions in the presence of $CdCl_2$, in contrast to some of the other protein-salt systems showing reentrant condensation[224]. In these cases, the liquid–liquid phase separation region is found as a closed area within the second regime.

We emphasize that the observed protein condensation is not caused by a change of the protein structure induced by $CdCl_2$. We have performed Fourier transform infrared (FTIR) and circular dichroism spectroscopy for a broad protein and salt concentration range as presented in Figure 4.3. Both techniques indicate no significant change on the secondary structure of the protein. Time-dependent FTIR spectra of the amide I and II bands also confirm that no significant structural change occurs during the entire crystallization process (S.I., Figure B.1a).

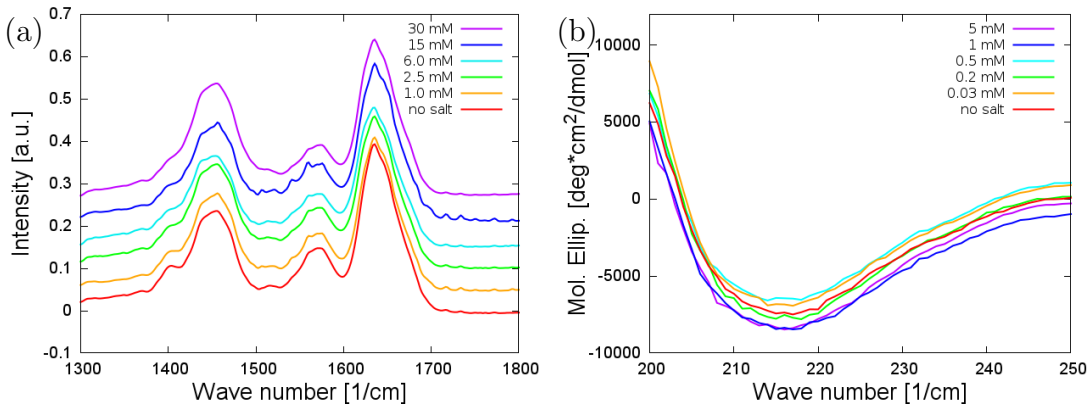

Figure 4.3: *(a) FTIR results (amide I and II) for a BLG concentration of 20 mg/mL and $CdCl_2$ concentrations between 0 and 30 mM (covering all three regimes). (b) Circular dichroism spectra for samples with a BLG concentration of 0.2 mg/mL and $CdCl_2$ concentrations between 0 and 5 mM.*

4.3.2 Crystallization Followed by Optical Microscopy

Crystallization of BLG in the presence of $CdCl_2$ was observed within a wide range of the above shown phase diagram. Slightly below and above c^*, no indications for a multistep process were observed: Crystals nucleate directly from a solution and grow larger without any visible clustering or liquid–liquid phase separation (Figure 4.4a-c). The typical shape of the crystals is needle-like under these conditions. In contrast, in the transition zone of *pseudo-*c^{**} (Figure 4.4d-f), protein aggregates form quickly after sample preparation and the solution observed by optical microscopy becomes more turbid. During crystallization, the turbid solution becomes gradually clearer, indicating that the aggregates are consumed. Removing the aggregates by centrifuging can significantly reduce the number of crystals, as also shown in other protein systems [181].

We further analyze the time dependence of the number and size of crystals. The visible area A of crystals in the microscopy pictures was determined by the open source program *ImageJ* [167]. The average crystal length can be estimated as $L \approx \sqrt{A/N}$, where N is the number of crystals in the respective image. Figure 4.4g shows the evolution of L and N with time. In the beginning, the number of crystals increases linearly and then saturates. After the visible crystals stop increasing in number, they still significantly increase in size. The increase in L is slow until 80 min, then faster growth starts and eventually saturates due to depletion of protein. The effect of centrifuging on the number of crystals and the acceleration of the growth rate after 80 min suggest that nucleation is closely related to the amount of protein portioned in aggregates formed in the early stage of sample preparation.

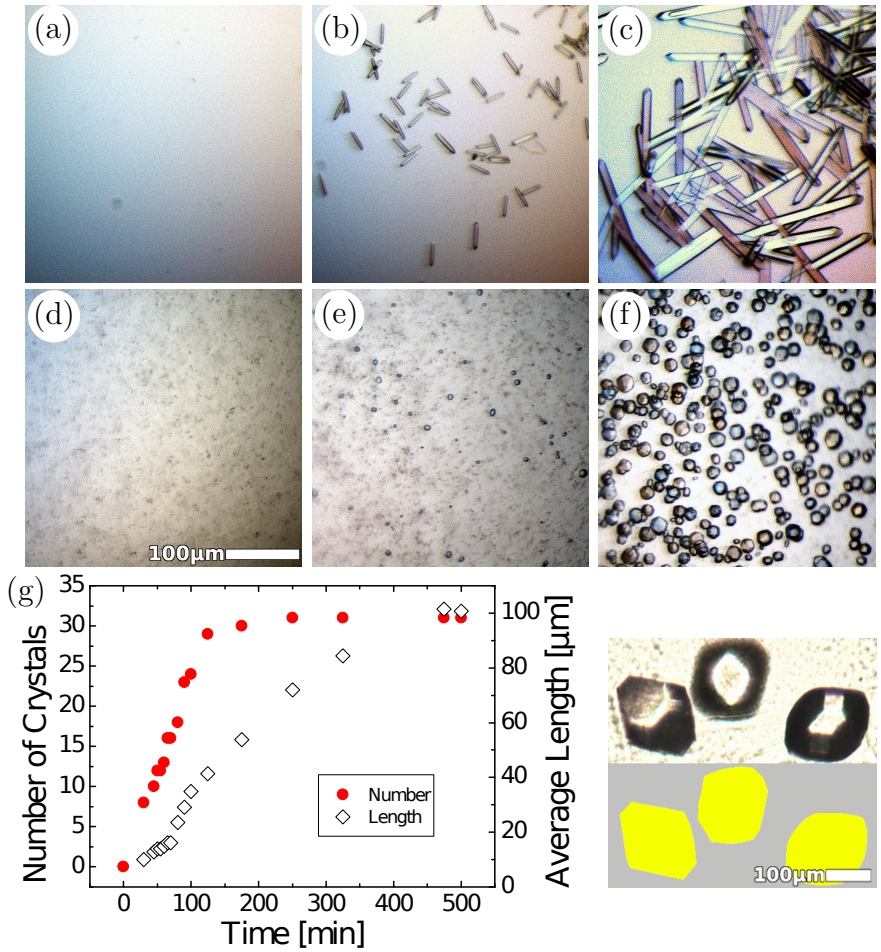

Figure 4.4: *Real-time observation of crystal growth by optical microscopy. (a-c): Crystallization of a clear sample slightly below c* (6.5 mg/mL BLG with 0.7 mM $CdCl_2$), directly after preparation (a), after 30 min (b) and after several days (c). (d-f): Nonclassical crystallization in the pseudo-c** zone (3.3 mg/mL BLG with 7 mM $CdCl_2$) 20 min (d), 30 min (e), and 120 min after preparation (f). (g): Quantitative analysis of a nonclassical crystallization process for a sample within pseudo-c** (20 mg/mL BLG with 15 mM $CdCl_2$, see also video in S.I.). The number of crystals N (filled red circles, left axis) and the average length $L = \sqrt{A/N}$ (open diamonds, right axis) are plotted as a function of time. Right: Evaluation of the area A of crystals.*

4.3.3 Crystal Growth Followed by Real-Time SAXS

Because of the fast crystallization process and limited resolution of optical microscopy, it is impossible to deduce the role of protein aggregates on the nucleation process by this method. For this purpose, we employ real-time SAXS to follow the early stage of crystallization. Figure 5.5 shows time-resolved SAXS data for crystallization slightly below c^* (Figure 5.5a) and within the transition zone of *pseudo-c*** (Figure 5.5c). By dividing each curve by the first one, we monitor the

evolution of the samples. This is shown in the bottom projections of the 3D-plots. In Figure 5.5b,d, $I(q,t)$ and $I(q,t)/I(q,t=0)$ are plotted at selective times for further discussion. For comparison, the form factor of BLG is also shown (blue dotted line).

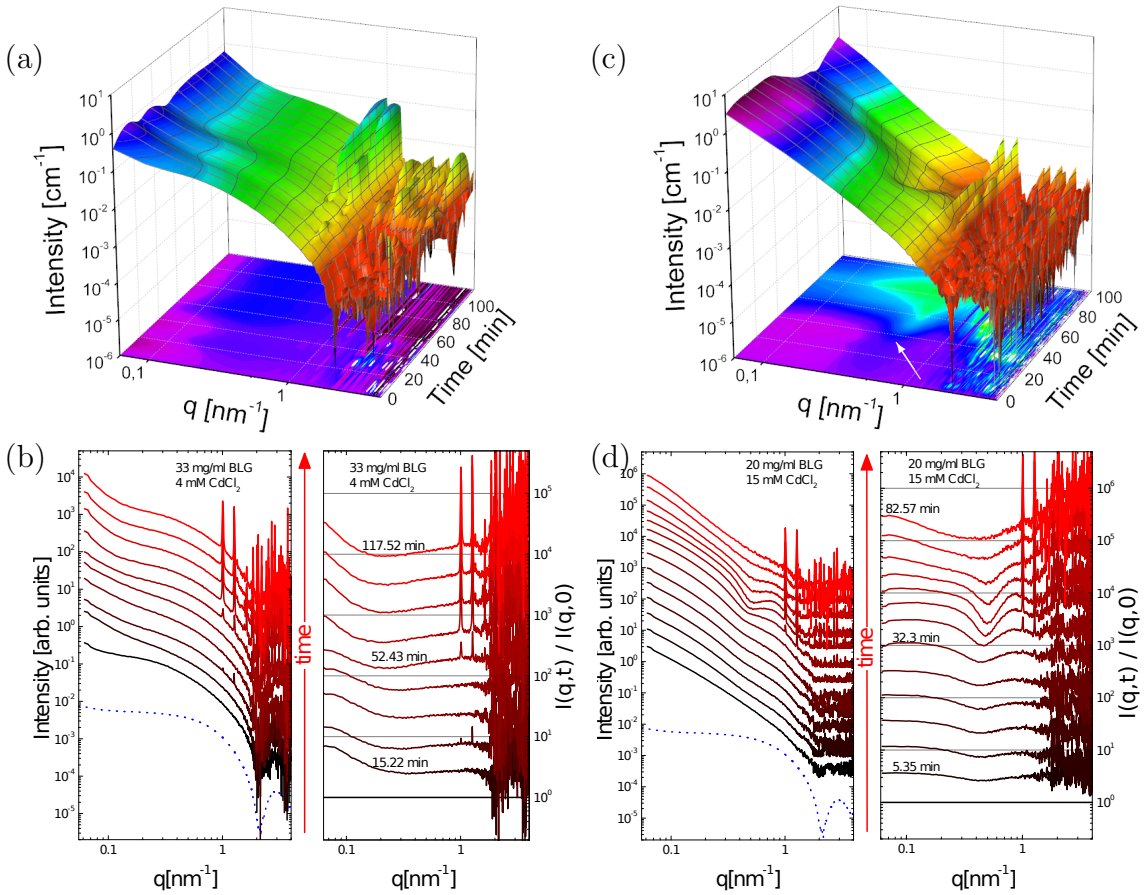

Figure 4.5: *Real-time SAXS curves for 33 mg/mL BLG with 4 mM (a-b), i.e. slightly below c*, and 20 mg/mL BLG with 15 mM $CdCl_2$ (c-d), i.e. within pseudo-c**. (a,c) 3D surface illustration of $I(q,t)$. The bottom projections are created from all curves divided by the first one. The arrow in (c) points to an additional structure forming before crystallization. (b,d) Single $I(q,t)$ and $I(q,t)/I(q,t=0)$ curves at selective times multiplied by constant factors for better visibility. The blue dotted line shows the form factor of BLG at low concentrations.*

Below c^*, the scattering curves change gradually with time. After an induction time (20 min in the example shown in Figure 5.5a-b), Bragg peaks appear and grow with time. Less pronounced Bragg peaks at later times are caused by crystals sedimenting out of the beam position. This is occasionally observed close to c^*, where the samples are less viscous than within the *pseudo-*c^{**} zone. Apart from that, the underlying shape of the curves barely changes with time. No indication of states other than solution and crystal was found. Similar behavior

was found for other samples slightly below and above c^*. Hence, close to c^*, the crystallization process is consistent with the observation by optical microscopy and can be explained by classical nucleation theory.

Importantly, within the region of *pseudo-c***, $I(q,t)$ follows a power law of q^{-2} at low q directly after preparation (the change of the exponent is shown in Figure B.2 in S.I.), deviating strongly from the form factor of BLG under physiological conditions (blue dotted line in Figure 5.5b,d). This is consistent with the formation of aggregates after sample preparation[63,99]. The data were analyzed by normalizing the time-dependent $I(q,t)$ by $I(q,t=0)$, which minimizes the effect of the form factor and the aggregate formation in the system and emphasizes the time-dependent relative changes. With increasing time, a broad maximum at $0.7-1.4\,\mathrm{nm}^{-1}$ appears (Figure 5.5c-d). Since scattering probes molecular and solution structures directly and thus is sensitive to changes of particle correlations, the time-dependent change of this broad peak can be interpreted as the evolution of a new structural feature within the larger aggregates at length scales of about 10 nm. Over time, this broad peak increases in intensity until Bragg peaks start to be observed. The two most pronounced Bragg peaks located at 1.01 and $1.27\,\mathrm{nm}^{-1}$ overlap with the broad maximum. The intensity of the Bragg peaks increases with time, whereas the broad peak shrinks and eventually disappears. Similar crystallization behavior was also observed for other samples within the region of *pseudo-c*** (cf. Figure 4.2).

To quantify the relationship between the aggregates (broad peak) and crystals (Bragg peaks) as a function of time, we use the concept of the degree of crystallinity known from semicrystalline polymers[185] for further data analysis: The intensity at the minimum (at around $0.5\,\mathrm{nm}^{-1}$) was subtracted to set I_{min} to zero. Then, the broad peak in $I(q,t)/I(q,t=0)$ was fitted by a broad scaled Gaussian, and the remaining Bragg peaks were fitted by two further (sharp) Gaussian functions (Figure 4.6a). For different times, the enveloped area of the broad region, A_{interm}, and the area of the Bragg peaks, A_{Bragg}, were followed over time, monitoring the crystallization kinetics.

Figure 4.6b shows A_{interm} (red spheres) and A_{Bragg} (green stars) as a function of time. The formation of the intermediate phase starts from the very first measurements on, increasing with time up to a maximum after approximately 42 min. Afterwards, A_{interm} decays gradually to zero. A_{Bragg}, on the other hand, is zero in the beginning and becomes visible after 30 min. Interestingly, the increase with time is nontrivial: After reaching a plateau around 50-80 min, A_{Bragg} increases dramatically. For further understanding, we show the first time derivative of A_{Bragg} which represents the crystallization rate (blue dashed line in Figure 4.6b, right axis), calculated at each point in time from the average of the incoming and outgoing slope. Interestingly, dA_{Bragg}/dt has a local maximum corresponding to the maximum of A_{interm}, suggesting that the crystallization rate is proportional to the amount of protein in these aggregates. Recalling the observation in Figure 4.4g, the early stage of crystallization can be interpreted as dominated by the nucleation events. This suggests that nucleation occurs within the intermediate

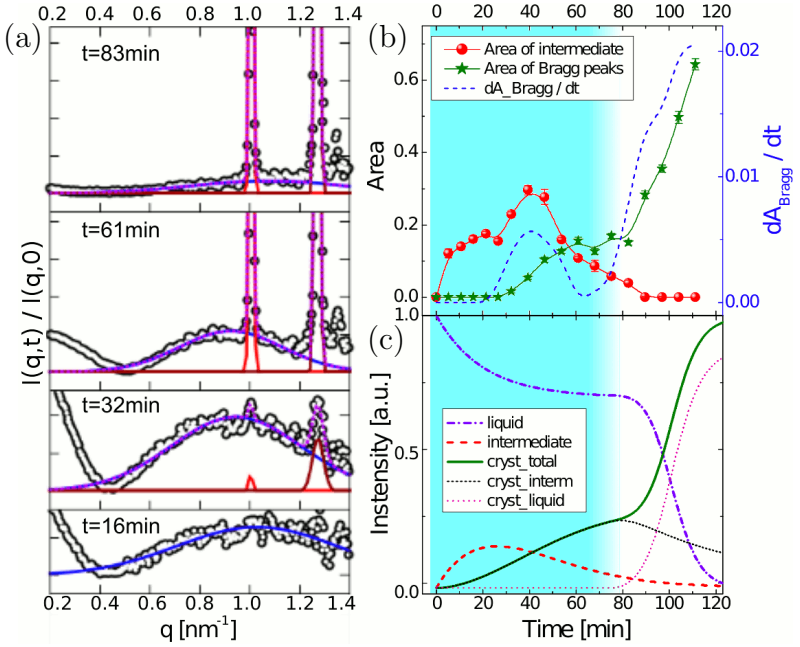

Figure 4.6: *Analysis of the time-dependent data. (a) For data analysis, the areas of the intermediate and crystalline part of $I(q,t)/I(q,t=0)$ were evaluated. (b) Red spheres (left axis) show $A_{interm}(t)$, the integral of the broad Gaussian function, connected to the intermediate state from 0.4 to 1.3 nm^{-1}. Green stars (left axis) show $A_{Bragg}(t)$, the integral of the two Bragg peaks in this area. The blue dashed line (right axis) shows the time derivative of $A_{Bragg}(t)$, the crystallization rate. The growth rate has a local maximum at the total maximum of $A_{interm}(t)$. (c) A rate equation model based on physically meaningful processes reproduces the experimental kinetic features. The light-cyan area marks the nucleation-dominated regime and is followed by the growth-dominated regime (no sharp transition).*

phase. The strong increase in A_{Bragg} after 80 min can be related to the growth-dominated stage as observed in Figure 4.4g. The same behavior could be observed for other conditions within the region of *pseudo-c*** followed by real-time SAXS. An analysis for a sample with 33 mg/mL BLG with 17 mM $CdCl_2$ can be found in the S.I., Figures B.3-B.5. In contrast, all tested conditions close to c^* show no signs of the existence of an intermediate phase between the initial solution and the final crystalline state.

4.3.4 Discussion: Parallel One-Step Processes versus Two-Step Process

Based on the above presented real-time results from both optical microscopy and SAXS, we suggest the following crystallization mechanism: In the first step, protein aggregates form as an intermediate and crystals nucleate from the aggregate precursor after a short induction time. The nucleation rate is thus proportional to

the amount of the intermediate phase; as expected for such a two-step nucleation, a local maximum in the crystallization rate (and hence a point of inflection in A_{Bragg}) occurs when A_{interm} is at its maximum. Crystal growth is slow, possibly due to the slow dynamics of molecules within the aggregates, which makes it difficult for proteins to diffuse to the growth front. However, such slow dynamics may be sufficient for nucleation as less molecules are involved in this process. We note that if the intermediate is in a kinetically arrested state (even slower dynamics), nucleation can be strongly impeded [202].

In the second step, the intermediate phase is consumed by nucleation and crystal growth, and the crystals grow rapidly by consuming free protein from the dilute phase. Proteins in the dilute phase (also in small aggregates as indicated from SAXS profiles and earlier studies [221]) can be easily incorporated into the crystal lattice. Importantly, the nucleation rate in the second step is significantly reduced, which further demonstrates the likely role of the protein aggregates as the precursor of a two-step nucleation process. All observations are explained very naturally using the two-step nucleation pathway based on clearly defined physical concepts (Figure 4.1b).

The proposed two-step nucleation mechanism can be further supported by using a rate equation model. Before looking into different quantitative models of the nucleation process, we note three qualitative model-free aspects directly from the data: (a) The intermediate forms very fast, while the initial crystal nucleation and growth are much slower. (b) Given the fast formation of intermediate, the disappearance of the intermediate has implications for its nature. At the time of disappearance, a considerable concentration in the liquid phase is still present, since the crystal growth is taking place even later. Thus, the intermediate is not in equilibrium with the liquid phase, but rather supplied by an excess concentration in the supersaturated liquid. (c) The plateau in the crystal formation indicates that the initial crystal growth is much slower than the final crystal growth.

One possible scenario involving an intermediate and crystals is a picture of two parallel one-step processes as sketched in Figure 4.1a. The intermediate would form from the supersaturated solution, and subsequently dissolve again, once the liquid phase is depleted by crystal nucleation and growth. While this scenario is consistent with observation (a) and (b) above, it cannot explain the plateau (c). Given the fast kinetics of the formation of the intermediate (a), the concentration in the liquid stays rather constant at the saturation concentration (b). Thus, once crystals are present, crystal growth should occur with a similar steeply increasing signature as in the late growth state, instead of being stagnant for some period. Thus, the parallel picture fails to explain an essential aspect observed in the experiment.

We note that it may be possible to introduce *ad hoc* technical modifications of the one-step parallel scenario to force-fit the data. Note that these modifications necessarily would violate a picture of a one-step process along classical nucleation theory, and rather suggest an effective multistep process. We also note that the existence of heterogeneous nucleation or the combination of heterogeneous and

homogeneous nucleation processes might lead to an overall growth kinetics similar to two or multiple parallel nucleation process, which again will not explain the plateau. Furthermore, these modifications would rely on a certain combination of parallel one-step processes, which makes the consequential two-step mechanism more feasible to explain the experimental results.

The second scenario is the picture of a two-step process of crystal nucleation from a previously formed intermediate as represented by Figure 4.1b. As shown in Figure 4.6c, a related model of rate equations based on processes with a clear physical meaning reproduces all features with very good agreement with the data. In the following, the system is described in terms of the mass fractions of protein in the liquid solution (L), in the intermediate phase (I) and within crystals in the solution (C_L) and intermediate phase (C_I); $(x)_+$ equals x for positive x and 0 otherwise.

First, the intermediate forms from the solution until the liquid solution reaches a stable mass fraction L_0: $\Delta_I = k_I(L-L_0)_+$. Second, crystals nucleate slowly from the intermediate: $\Delta_n = k_n I$. Third, the crystal growth within the intermediate depends on the amount of crystals and intermediate: $\Delta_{gI} = k_{gI} I C_I$. Forth, once the intermediate phase is consumed and falls below a certain volume per crystal, the crystals become exposed to the liquid phase: $\Delta_t = k_t C_I(\alpha_I C_I - I)_+$. α_I is the ratio of the critical intermediate volume per crystal volume. Finally, once crystals emerged into the liquid phase, the crystal growth in the liquid phase consumes the free protein molecules: $\Delta_{gL} = k_{gL} L C_L$.

Consequently, the rate equations read

$$\partial_t L = -\Delta_I - \Delta_{gL} \tag{4.1}$$

$$\partial_t I = \Delta_I - \Delta_n - \Delta_{gI} \tag{4.2}$$

$$\partial_t C_I = \Delta_n - \Delta_t + \Delta_{gI} \tag{4.3}$$

$$\partial_t C_L = \Delta_e + \Delta_{gL} \tag{4.4}$$

with the initial conditions $L(0) = 1$, $I(0) = 0$, $C_I(0) = 0$, and $C_L(0) = 0$. The temporal evolution has been calculated with the Euler method using the following choices of model parameter: $k_I = 0.05\,\mathrm{min}^{-1}$, $k_n = 0.02\,\mathrm{min}^{-1}$, $k_{gI} = 0.1\,\mathrm{min}^{-1}$, $k_{gL} = 0.2\,\mathrm{min}^{-1}$, $k_e = 1.0\,\mathrm{min}^{-1}$, $L_0 = 0.7$, and $\alpha_I = 0.2$. Note that the absolute values of the rates are consistent with the physical picture from this work and all parameters take physically reasonable values. We emphasize that the general qualitative result is robust against slight changes of the absolute values.

4.4 Conclusion

In summary, our real-time SAXS and optical microscopy study of protein BLG crystallization in the presence of $CdCl_2$ provides evidence for a two-step nucleation mechanism, i.e. protein aggregates form first as an intermediate and crystals nucleate within these aggregate precursors after an induction time. The nucleation

rate is thus proportional to the amount of the intermediate phase, consistent with the appearance of a local maximum in the crystallization rate at the maximum quantity of the intermediate. However, the crystal growth rate is low, which might be due to the low mobility of proteins within the aggregates. This step lasts until the intermediate phase is consumed by crystallization, then crystal growth becomes faster as proteins in the dilute phase can diffuse to and incorporate into the growth front directly. In contrast, the nucleation rate is significantly reduced, which further demonstrates the role of the protein aggregates as the precursor of a two-step nucleation process. The experimental observations on the nonclassical growth kinetics have been interpreted using a rate equation model. In addition, these measurements demonstrate a noninvasive method to study *in situ* and in real-time the kinetics of nonclassical growth processes on nanometer length scales.

The authors thank the DFG for financial support and H.-J. Schöpe and M. Oettel for valuable discussions and comments. A.S. acknowledges a fellowship of the Landesgraduiertenförderung. A.S. thanks V. Niemann for the assistance in CD measurements, and T. Stehle (IFIB, Universität Tübingen) for sharing their lab resources.

Chapter 5

Publication C. *On the Question of Two-Step Nucleation in Protein Crystallization*

Faraday Discuss. ***179*** *(2015), 41–58.*

Andrea Sauter, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Artem Feoktystov, Robert M. J. Jacobs and Frank Schreiber

Contributions	
Research design	*AS, FS, FZ*
Experiments	*AS, FZ*
Technical Assistance	*AF, GL*
Rate equation model	*FRR*
Data analysis and Interpretation	*AS, FRR, FS, FZ, RMJJ*
Paper writing	*AS, FZ, FRR, FS*

We report a real-time study on protein crystallization in the presence of multivalent salts using small angle X-ray scattering (SAXS) and optical microscopy, particularly focusing on the nucleation mechanism as well as the role of the metastable intermediate phase (MIP). Using bovine beta-lactoglobulin as a model system in the presence of the divalent salt $CdCl_2$, we have monitored the early stage of crystallization kinetics which demonstrates a two-step nucleation mechanism: protein aggregates form a MIP, followed by the nucleation of crystals within the MIP. Here we focus on characterizing and tuning the structure of the MIP using salt and related effects on the two-step nucleation kinetics. The results suggest that increasing the salt concentration near the transition zone $pseudo - c^{**}$ enhances the driving force for both MIPs and crystal nucleation leading to slow crystallization. The structural evolution of the MIP and its effect on subsequent nucleation is discussed based on the growth kinetics. The observed kinetics can be well described using a rate-equation model based on a clear physical two-step picture. This real-time study not only provides evidence for a two-step nucleation process

for protein crystallization, but also elucidates the role and the structural signature of the MIPs in the nonclassical process of protein crystallization.

5.1 Introduction

Studies of the early stage of nucleation in various systems have revealed new features which cannot be explained using the classical nucleation theory [34,57,58,83,171,211]. A large body of experimental results supported by theory and simulations suggest that a metastable intermediate phase (MIP) exists before the final crystal structure is formed [20,30,64,65,79,100,137,138,161,164,168,181,196,198,228], i.e. the solutes in the supersaturated solution form in a first step either small clusters or a macroscopic dense liquid phase. In the second step, the nucleation occurs within the MIPs. This two-step nucleation mechanism has originally been proposed by ten Wolde and Frenkel for crystallization of a colloidal system with short-range attraction and near the critical point of the metastable liquid–liquid coexistence line [196]. The two-step nucleation mechanism can be considered as an example of Ostwald's step rule in the microscopic world [197]. Later, this concept has been widely used under various conditions [34,57,58,83,161,168,171,181,202,211,228].

While the two-step mechanism seems plausible for certain experiments, direct observation of such a process is not easy. Recently, direct visualization of the crystallization kinetics and pathways of nucleation in colloidal crystallization became possible and provided detailed information on the MIP and the transition in colloidal suspensions. Colloidal systems exhibit similar phase behavior as atomic and molecular systems, and their large particle sizes enable visualization on a single-particle level. Using this technique, Tan et al. studied the liquid-solid phase transition and observed the formation of a metastable precursor under their experimental conditions, regardless of the final state and the interaction potential [190]. Peng et al. studied the kinetics of a solid-solid phase transition using single-particle resolution video microscopy. They observed that the transition between two different solid states occurs via a two-step diffusive nucleation pathway involving liquid nuclei [141]. This pathway is favored in comparison with the one-step nucleation because the energy of the solid/liquid interface is lower than that between the solid phases.

While these experimental observations on colloidal systems demonstrate that the two-step nucleation follows Ostwald's step rule for simple liquids [88,107,161,198], the application of this concept to other systems, in particular the protein crystallization, is still challenging. The dimensions of proteins on the nanometer scale limit the applicability of optical methods for the study of the MIP formation. Because of the larger size and slow dynamics of colloids, the microstructural arrangement of colloidal particles relaxes typically in a time scale of seconds, which leads to various non-equilibrium phenomena in these systems. Moreover, the interaction potentials in these colloidal systems are isotropic, whereas the effective protein-protein interactions are often non-isotropic involving hydrophobic and

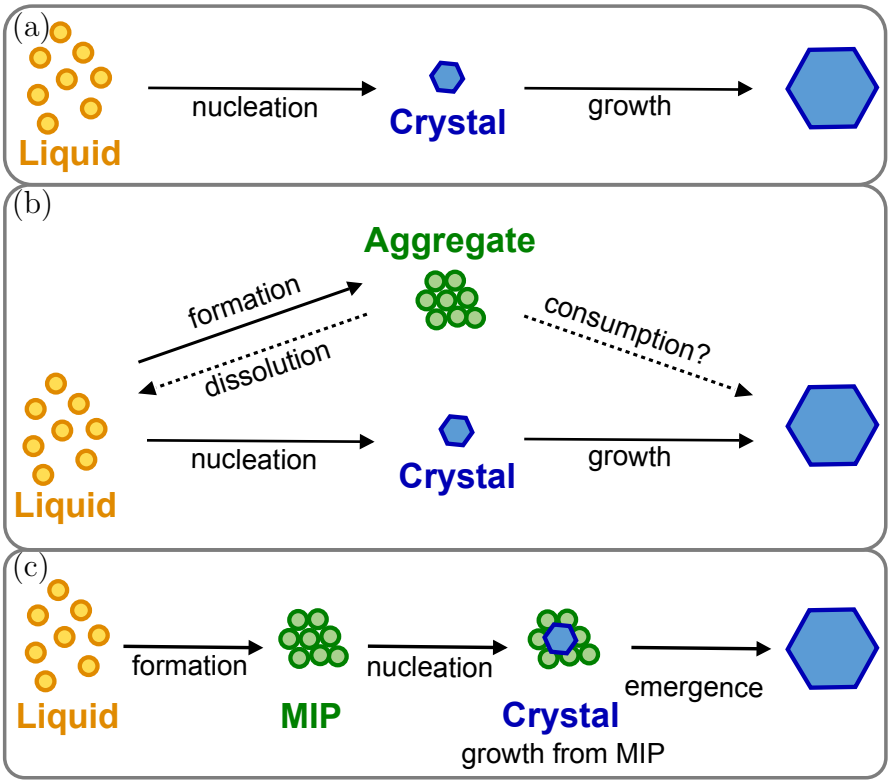

Figure 5.1: *Illustrations of different pathways of crystal nucleation and growth. (a) Classical one-step pathway. (b) Parallel process of aggregate formation and crystal nucleation. (c) Multi-step process with precursor (MIP) for nucleation.*

electrostatic patches as well as ion-bridges. The quantitative description remains poorly understood [143,173,225].

For these systems, in which the direct visualization is not possible, an other method has to be developed to characterize the MIP and the nucleation and growth kinetics. Here, we argue that the two-step nucleation can be distinguished from the classical one-step process by following the overall crystallization kinetics. When a MIP exists, particular care should be paid for distinguishing the sequential and the parallel pathway. The sequential pathway corresponds to the real two-step nucleation, in which the crystals nucleate from the MIP. The parallel pathway consists of the parallel formation of MIP and crystals in two one-step nucleation events from the liquid. The different pathways are illustrated in Figure 5.1.

For a one-step nucleation and growth mechanism with or without MIPs, the nucleation and growth are mainly determined by the supersaturated initial solution. The overall growth kinetics is expected to show an increase first and to saturate later. By contrast, for a two-step mechanism, nucleation may occur in both the MIPs and the initial supersaturated solution, and the nucleation rates are different due to the different interfacial energies. Furthermore, the subsequent crystal growth rates are different too. Therefore, a characteristic two-step growth kinetics

is expected, dominated by MIP and the initial solution, respectively. More details will be discussed based on the rate equation model in this paper. Indeed, in our recent study using time-resolved SAXS and optical microscopy, we have identified the kinetic signature of two-step nucleation in crystallization of the protein β-lactoglobulin (BLG) in the presence of $CdCl_2$ [160]. The first step of kinetics is dominated by the two-step nucleation and growth within the MIP, i.e. protein aggregates as a MIP are formed, followed by the nucleation of crystals within the MIP. During this period, the number of crystals increases with time, but the growth rate is slow due to the low mobility of surrounding aggregates. The second step of kinetics is dominated by the growth in the dilute phase: the consumption of the MIP leads to the exposure of the crystals to the free molecules in the dilute phase, which accelerates the crystal growth. In this work, we focus on the effect of the structural property of MIP on the nucleation and growth of the two-step nucleation mechanism.

5.2 Experimental

5.2.1 Materials and Sample Preparation

The globular protein β-lactoglobulin (BLG) from bovine milk (product no. L3908), $CdCl_2$ (202908) was purchased from Sigma-Aldrich. For sample preparation, appropriate amounts of salt stock solution, millipore water and protein stock solution were mixed. Stock solutions were prepared by dissolving the salt or protein powder in deionized (18.2 MΩ) and degassed millipore water. The protein concentration of stock solutions was determined by UV absorption measurements using an extinction coefficient of $0.96\,\mathrm{l}\cdot\,\mathrm{g}^{-1}\cdot\,\mathrm{cm}^{-1}$ at a wavelength of 278 nm [182]. All samples in this work were prepared without additional buffer since buffers can affect the phase behavior of proteins and the solubility of salts. The pH of the solutions was monitored using a Seven Easy pH instrument from Mettler Toledo. The pH values for all experimental conditions were above the $pI = 5.2$ of BLG [44]. Therefore, cation binding is the main driving force of charge inversion instead of pH [44,151,226]. All experiments were performed at room temperature of $293 \pm 1\,\mathrm{K}$.

5.2.2 Optical Microscopy

Time-dependent protein crystallization was followed by the transmission optical microscope AxioScope.A1 from Zeiss. The protein stock solution was filtered (pore size 100 µm) in advance. Samples were prepared using a micro-batch setup with two hydrophobically coated glass slides sealed by silicone (sample thickness ≈ 250-300 µm). Images were taken by an included camera AxioCam ICc5.

5.2.3 Small-Angle X-Ray and Neutron Scattering (SAXS and SANS)

Small angle X-ray scattering (SAXS) measurements were performed at the ESRF, Grenoble, France at the beamline ID02. Different energies of 16047 eV and 12460 eV were used for two different beamtimes, and the sample-to-detector distance was 2 m with an accessed q-range of 0.06 to 4.3 nm^{-1} or 0.04 to 3.9 nm^{-1}, respectively. Ex-situ measurements were performed using a flow capillary cell. For real-time measurements, samples were measured using quartz capillaries in a vertical capillary holder that were quickly loaded and transferred to the sample station after sample preparation. Measurements started about 2-3 min after mixing and took place every couple of minutes during the whole crystallization process. The beam position in the sample was shifted after each exposure (duration 0.05 s) to avoid radiation damage. For further details on the beamline, calibration and data collection see Ref. [129].

SANS measurements were carried out at KWS1 at FRMII, Munich, Germany. The applied sample-to-detector distances were 1.5 and 8 m which covers a q-range from 0.04 nm^{-1} to 3.1 nm^{-1} at a wavelength of 7 Å ($\Delta\lambda/\lambda = 10\,\%$). Protein-salt solutions in D_2O were filled in rectangular quartz cells with a path-length of 2 mm. The beam size on the sample was 6 mm x 12 mm. Plexiglas was used as secondary standard to calibrate the absolute scattering intensity. The data correction and absolute intensity calibration were obtained using the software QtiKWS[144].

5.3 Results

5.3.1 Experimental Phase Diagram of BLG with $CdCl_2$ or $ZnCl_2$

We first describe the experimental phase diagram of our system which provides the basis for the following kinetic studies on protein crystallization. Our studies of globular proteins in solutions containing multivalent metal ions have revealed complex phase behavior including reentrant condensation (RC), metastable liquid–liquid phase separation (LLPS) and crystallization[86,151,183,220,222,224,226,227]. A similar experimental phase diagram, like the one shown in Figure 5.2, has been observed for several proteins in solution in the presence of trivalent metal ions. The physical mechanism of this RC behavior is due to the effective charge inversion of proteins and a cation-mediated attraction, presumably by intermolecular bridges of multivalent cations[152,227]. With an isoelectric point below 7, the proteins used in our work are acidic at neutral pH. At low salt concentrations, proteins carry negative net charges, and the electrostatic repulsion stabilizes the solution. Adding trivalent metal ions into the solution, the binding of metal ions to the carboxyl groups on the protein surface reduces the effective net charge. Above a certain salt concentration c^*, electrostatic repulsion is not strong enough to

balance the attractive potential, and samples phase separate and become turbid ("regime II"). The interesting observation of this system is that with further increasing salt concentration, the continuous binding of metal ions to the protein surface leads to a charge inversion, which again establishes the long range electrostatic repulsion. Therefore above a second boundary (experimentally rather broad), c^{**}, the solutions become completely clear again. The charge inversion and the effective attraction mediated via multivalent metal ions have been further investigated by experiments, simulations and theoretical studies [5,86,92,151,152].

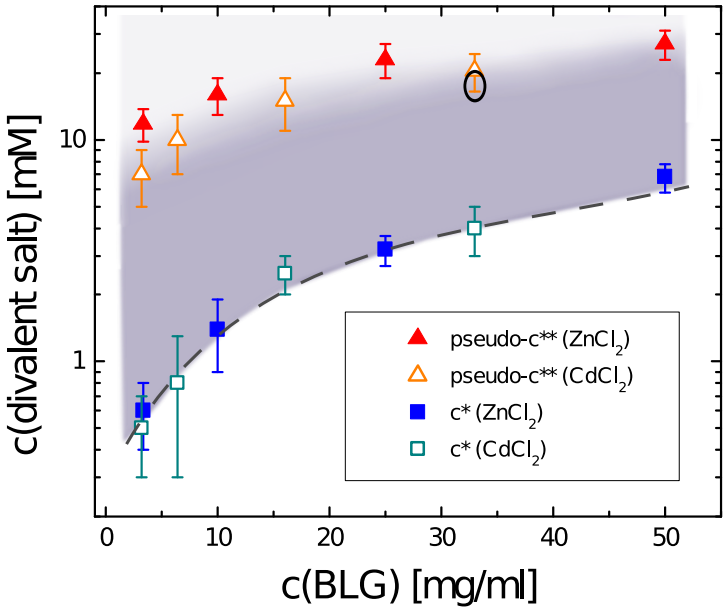

Figure 5.2: *Experimental phase diagram of BLG with $ZnCl_2$ and $CdCl_2$ [158,160]. Both salts share the similar c^*, below and above which the protein solutions are clear and turbid, respectively. Further increasing salt concentration, samples become gradually less turbid, but not completely clear again, and this transition zone is denoted as $pseudo - c^{**}$. The black ellipse marks the conditions mainly used in this work (33 mg/mL BLG with 15-20 mM $CdCl_2$).*

Previous studies on β-lactoglobulin (BLG) systems with divalent salts $ZnCl_2$ and $CdCl_2$ showed a similar experimental phase behavior (Figure 5.2) [158,160]. For these divalent salts, the samples above a certain salt concentration become gradually less turbid, but not completely clear again, and this transition zone is denoted as $pseudo - c^{**}$. Both boundaries induced by $CdCl_2$ and $ZnCl_2$ are remarkably similar as shown in Figure 5.2. In comparison to the trivalent salt YCl_3, both transitions are shifted towards higher salt concentrations [158]. Although the reentrant effect is not complete, a charge inversion with increasing divalent salt concentration has been observed in both cases (S.I., Figure C.1). Note that the phase behavior shown in Figure 5.2 was observed for BLG only, but not for bovine or human serum albumin (data not shown), which suggests a more specific interaction between these divalent ions and BLG [158].

We emphasize that the observed protein condensation is not caused by changing of the protein structure induced by $CdCl_2$, $ZnCl_2$ or other multivalent salts, as demonstrated in previous work using Fourier transform infrared (FTIR) and circular dichroism spectroscopy for a broad protein and salt concentration range [80,160]. Both techniques indicate no significant change on the secondary structure of the protein. Moreover, the successful growth of high-quality crystals and fine structural analysis confirm that the proteins are still in their native state [158,227].

5.3.2 Crystallization Followed by Optical Microscopy

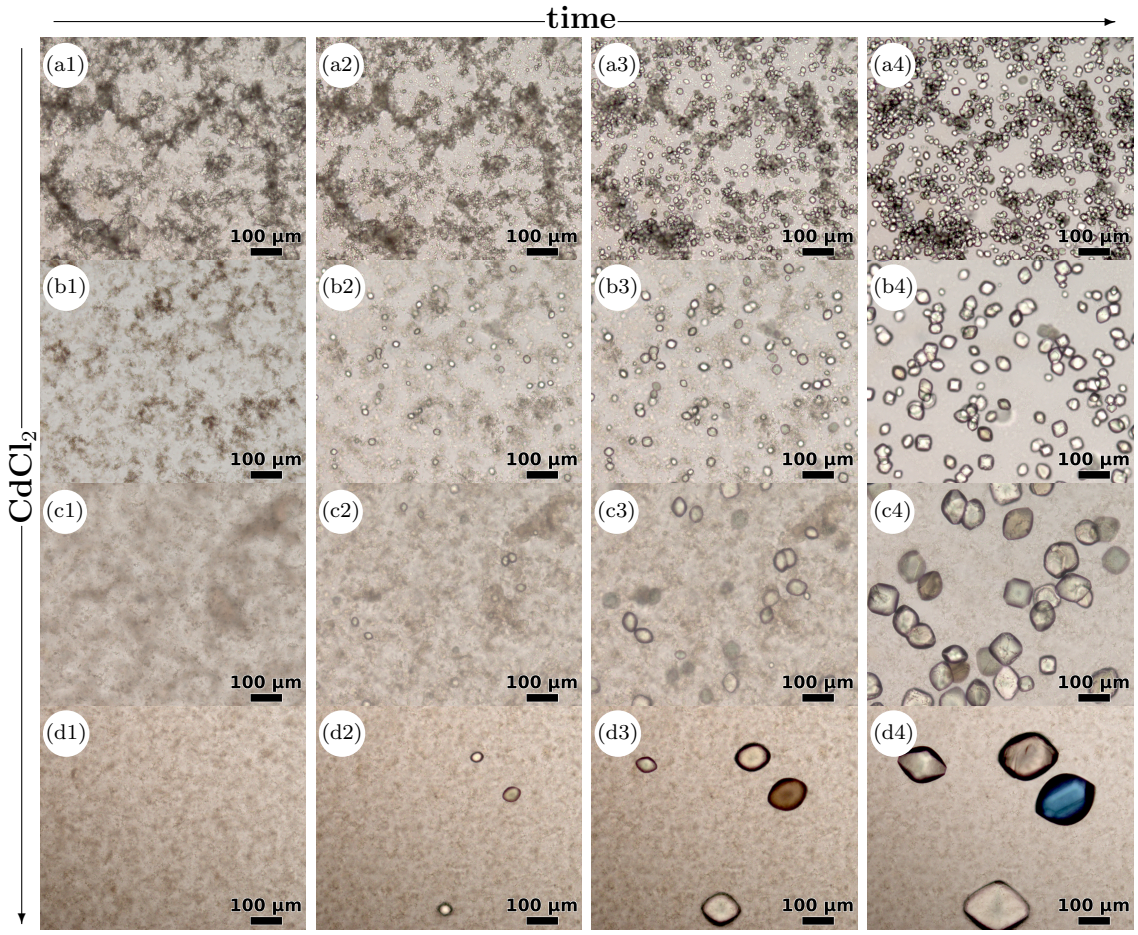

Figure 5.3: *Image series of protein crystallization for samples with a BLG concentration of 33 mg/mL and $CdCl_2$ at different time: (a) 17 mM, (b) 18 mM, (c) 19 mM and (d) 20 mM (increasingly less turbid). (a1) 0, (a2) 50 min (a3) 75 min (a4) 100 min after preparation. (b1) 0 (b2) 75 min (b3) 100 min (b4) 3 h. (c1) 0, (c2) 100 min, (c3) 3 h, (c4) 6 h. (d1) 0, (d2) 3 h, (d3) 6 h and (d4) 10 h. Videos of the crystallization process in (a) and (d) can be found in the online S.I.*

Crystallization of BLG in the presence of $CdCl_2$ and $ZnCl_2$ was observed within a wide range of the above shown phase diagram (Figure 5.2) [158,160]. Slightly be-

low c^*, no indications for a multi-step process were observed: Crystals nucleate directly from a solution and grow larger without any visible aggregation or LLPS. In contrast, in the transition zone of *pseudo*-c^{**}, protein aggregates form quickly after sample preparation and the solution observed by optical microscopy becomes more turbid. In the case of $ZnCl_2$, the aggregates further relax into a liquid-like network before crystallization starts[158]. During crystallization, the turbid solution becomes gradually clearer, indicating that the aggregates are consumed. If the aggregates are removed by centrifugation, the number of crystals or the nucleation rate can be significantly reduced, as also shown in other protein systems[181]. The two different growth pathways have been observed for the same protein, BLG with a trivalent salt, YCl_3[158,227].

Within the $pseudo - c^{**}$ region, the crystallization of BLG in the presence of $CdCl_2$ follows a two-step pathway via aggregates[160]. Here, we study a series of samples with fixed protein concentration but varying salt concentration. Variation of the salt concentration changes the strength of the attractive potential, therefore the driving force of formation of both MIP and nucleation. Crossing $pseudo - c^{**}$ from low to high salt concentrations, the driving force is reduced. Figure 5.3 shows time series of images of crystallization for samples with 33 mg/mL BLG and $CdCl_2$ concentrations of 17, 18, 19 and 20 mM. Below 17 mM, the sample solution is deep in regime II, where massive amorphous aggregates form and precipitate quickly from the solution due to the strong attraction. Although crystallization can still be observed, only a small fraction of the amorphous aggregates can turn into the crystalline phase. Above 20 mM, solutions become nearly clear, and only small aggregates are visible under microscope. Crystallization in these solutions becomes extremely slow or do not crystallize at all.

From the images shown in Figure 5.3, one can see that with 17 mM salt, large aggregates are still formed, but the nucleation rates are also high. We emphasize that nearly all crystals are associated with the network. Nearly no crystals form in the dilute phase. In contrast to the high nucleation rate, the crystal growth period is short. After about 2 h, no visible change can be observed anymore, and the resulting system contains a large number of small crystals and most of the network or the amorphous aggregates has turned into the crystalline phase. Increasing the salt concentration by only 1 mM, to 18 mM, the overall phenomenological picture changes dramatically. Large aggregates are still visible, but not well connected to each other. The number of crystals is reduced but the final crystals are bigger. Further increasing the salt concentration, protein aggregates become smaller and the number of crystals is further reduced, but the crystal size is larger. In the end, the MIP is consumed by crystal growth and the solutions become clear. The averaged number of crystals normalized by the area of 1 mm^2 as a function of time is plotted in Figure 5.4 for three conditions. The number of crystals increases first with time, then saturates. The nucleation rates in the early stage are obtained from the slope of a linear fit, they are 1.44 ± 0.08, 0.32 ± 0.03 and 0.08 ± 0.01 min^{-1} for 18 mM, 19 mM and 20 mM $CdCl_2$, respectively. This decrease of nucleation rate is expected as the driving force is reduced with increasing salt concentration.

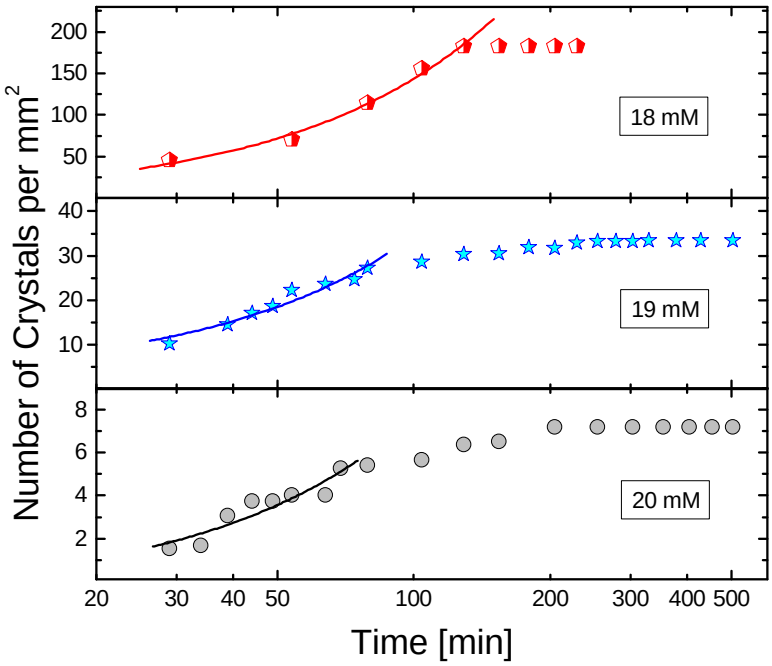

Figure 5.4: *Number of crystals as a function of time in a normalized area from Figure 5.3 for samples with 18 mM, 19 mM and 20 mM $CdCl_2$. For the 19 and 20 mM plots, three series of picture sets were analyzed and averaged for better statistics.*

5.3.3 Structure of MIP Revealed by SAXS and SANS

Due to the limited resolution of optical microscopy, SAXS and SANS were used for studying the structure and the role of the MIP on the crystallization process. We first show typical SAXS results in Figure 5.5a for samples with a low protein concentration of 6.5 mg/mL and $CdCl_2$ concentrations covering all three regimes. In regime I, the scattering curve (with 0.5 mM salt) is dominated by the form factor of the BLG dimer, which is consistent with the literature that BLG occurs predominantly as a dimer at room temperature and pH between 3.5 and 7.5 [207]. In regime II, with 1 mM salt, the high q part is still dominated by the form factor, but in the low q region, the increase of intensity indicates the formation of protein aggregates. In the third regime (12 - 90 mM, not all data are shown), a new feature forms shortly after preparation at $q \approx 0.7\,\mathrm{nm}^{-1}$ and a sharp peak occurs at $2.2\,\mathrm{nm}^{-1}$, as indicated by the arrows in this figure. A previous study of BLG in solution in the presence of YCl_3 has shown that this maximum corresponds to the monomer-monomer correlation due to the bridging effect [221]. Here, in the presence of $CdCl_2$, the peak is sharper. A possible explanation is the formation of a highly ordered fiber-like structure which gives such diffraction peak corresponding to the axial translation of the subunit (BLG monomer). Similar diffraction peaks have been observed in the solution scattering of F-actin [156].

Figure 5.5b presents SAXS data for samples under similar conditions as for the

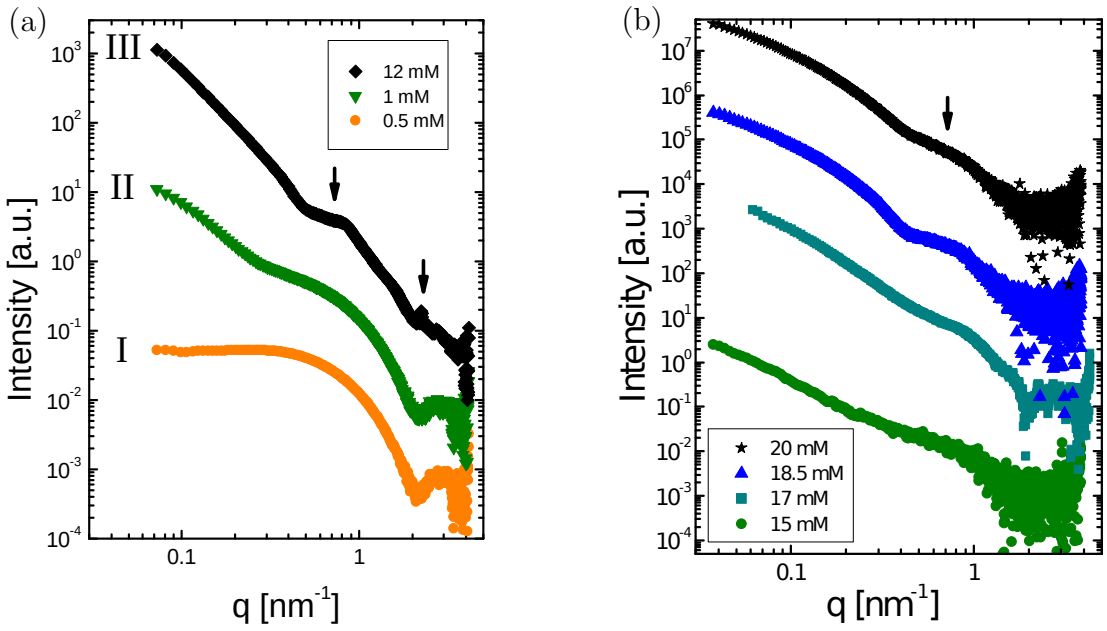

Figure 5.5: *(a) SAXS measurements at a protein concentration of 6.5 mg/mL with $CdCl_2$ concentrations covering all three regimes of RC (marked by I, II, III). (b) Selected SAXS profiles at the time when the broad peak is well developed for samples in regime III with 33 mg/mL BLG and 15, 17, 18.5 and 20 mM $CdCl_2$, respectively. Data are shifted upward for clarity.*

real-time optical microscopy study in Figure 5.3. The SAXS profiles were collected during the crystallization process at a time at which the broad peak at $q \approx 0.7\,\mathrm{nm}^{-1}$ has already developed, but Bragg peaks have not appeared yet. With 15 mM salt, the sample is located deep in regime II. While large aggregates are visible from optical microscopy (Figure 5.3a), SAXS profile shows a rather weak shoulder at $q \approx 0.7\,\mathrm{nm}^{-1}$, indicating less order within the large aggregates. Increasing salt concentration increases the ordering inside of the aggregates as the broad peak becomes more pronounced. The data set with 17 mM salt was collected at a different beamtime with higher resolution and also shows the monomer–monomer correlation peak at q around $2.2\,\mathrm{nm}^{-1}$.

We further measured a similar set of samples using SANS. In this case, the samples were prepared in D_2O instead of H_2O for better contrast. Neutron scattering causes no radiation damage to the proteins in solution, and most importantly, SANS measurements apply a large beam size (6 mm x 12 mm) which covers a much larger volume containing a reasonable number of crystals. This is in contrast to SAXS measurements with a beam size of 0.3 x 0.15 mm which contains a limited number of crystals in the case of crystallization in Figure 5.3c and d. The measurements were carried out for freshly prepared samples, 2 h and 6 h after preparation, respectively. The results shown in Figure 5.6 suggest that the samples behave very similar as in H_2O. First of all, in all cases, the fresh samples show the broad peak at $q \approx 0.7\,\mathrm{nm}^{-1}$ which has been identified as the characteristic

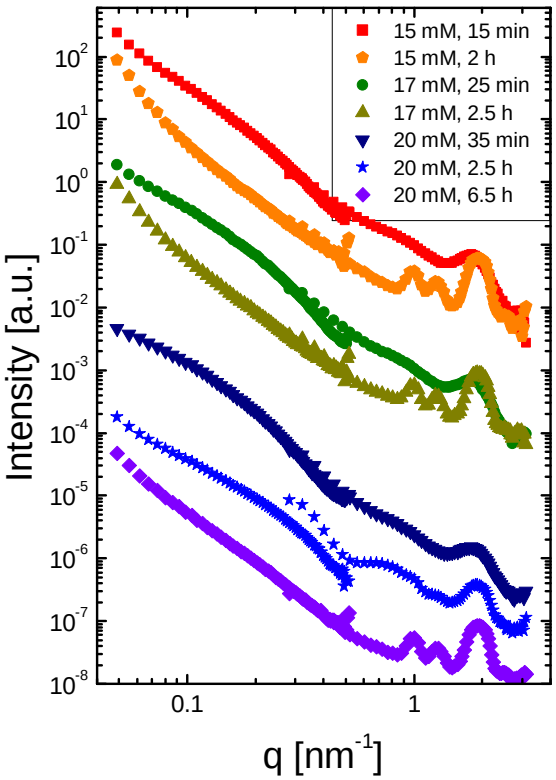

Figure 5.6: *BLG 33 mg/mL with 15, 17 and 20 mM $CdCl_2$ in D_2O measured by SANS. Data are shifted upward for clarity. The mismatch of the SANS curves (especially visible for 20 mM after 2.5 h) was caused by measurements at different detector positions together with the fast kinetics occurring in the sample.*

structural feature of the MIP. However, due to the low time resolution, the details of the development of this peak cannot be followed. Nevertheless, for the sample with 20 mM salt, it is visible that this broad peak develops and becomes more pronounced after 2 h. In addition to this broad peak, the monomer–monomer correlation peak at q around 2 nm^{-1} is pronounced in all cases which is in good agreement with the SAXS measurements. The slight shift of monomer–monomer peak to the low q value in SANS is due to the hydration effect [223]. Secondly, after certain time, smeared Bragg peaks appear for all samples. At the same time, the broad peak (MIP) reduces its intensity or completely disappears. Although the low resolution of SANS at the high-q region smears the Bragg peaks, their positions are consistent with the SAXS measurements.

Both SAXS and SANS measurements reveal the similar structural feature of MIP, i.e. the local ordering within the large protein aggregates characterized by a broad peak at $q \approx 0.7$ nm^{-1} and the monomer–monomer correlation peak at q around 2 nm^{-1}. As discussed in the following section, the broad peak is closely related to nucleation and crystal growth. It thus becomes the structure signature of MIP.

5.3.4 Crystallization Kinetics Followed by Real-Time SAXS

To extract information on the underlying crystallization process, we employed real-time SAXS measurements on the crystallization kinetics with high time and structural resolution. Figure 5.7 shows examples of time-resolved SAXS measurements for 33 mg/mL BLG with 17 (a) and 20 mM $CdCl_2$ (b) in 3D surface illustration and 2D projection, additional data for samples with 15 and 18.5 mM $CdCl_2$ were shown in S.I., Figure C.2. The bottom 2D projections are created by dividing all curves by the first one and therefore visualize the ongoing changes in the system with time. Selected $I(q,t)/I(q,t=0)$ curves are further presented in Figure 5.7c&d. The SAXS curves of both samples feature a strong increase at low q which hardly changes with time, indicating the presence of large aggregates consistent with the observation by optical microscopy. With increasing time, a broad peak located at $q \approx 0.7\,\mathrm{nm}^{-1}$ develops that has been assigned to the nucleation precursors (MIP)[160]. Once Bragg peaks appear, the most pronounced ones at 1.01 and $1.27\,\mathrm{nm}^{-1}$ overlap with the broad peak. The intensity of the Bragg peaks increases with time, while the broad peak shrinks. In the $I(q,t)/I(q,t=0)$ plots (Figure 5.7c&d), it is clear that the broad peak appears before the Bragg peaks and becomes stronger with time. At the end of the crystallization, the broad peak shrinks and eventually disappears (Figure 5.7a&c).

From the optical microscopy experiments (Figure 5.3) one observes that the MIP forms before crystallization starts and is consumed during crystal growth. From real-time SAXS measurements one can see that the typical broad peak for MIP follows the same development: it appears first and develops and once crystallization starts it reduces its intensity and eventually disappears. Based on these observations, we propose to use the relative change of the area of this peak (representative of MIP) and the two Bragg peaks which overlap with it to quantify the relationship between the MIP and the crystalline phase as a function of time. At this point, we use the concept of crystallinity from semi-crystalline polymer systems for further data analysis[185]. After subtraction of the intensity at the minimum, the broad peak in the $I(q,t)/I(q,t=0)$ curves was fitted by a scaled Gaussian function and the Bragg peaks by two further (sharp) Gaussians[160]. The crystallization kinetics can be followed by the enveloped area of the broad region, A_{interm}, and the area of the Bragg peaks, A_{Bragg}, as a function of time. This method is further illustrated in an animation that can be found in the online S.I. Figure 5.8a displays an example of such analysis for a sample with 33 mg/mL BLG and 17 mM $CdCl_2$. The development of MIP (A_{interm}) shows a maximum around 40 min, and the overall crystallinity (A_{Bragg}) has a plateau between 40 and 60 min, and then grows faster. Interestingly, the overall growth rate, i.e. the first derivative of A_{Bragg} on time, gives a maximum located also around 40 min, indicating that the overall crystal growth rate in the early stage strongly depends on the development of the MIP.

We have performed real-time SAXS measurements on all four salt conditions

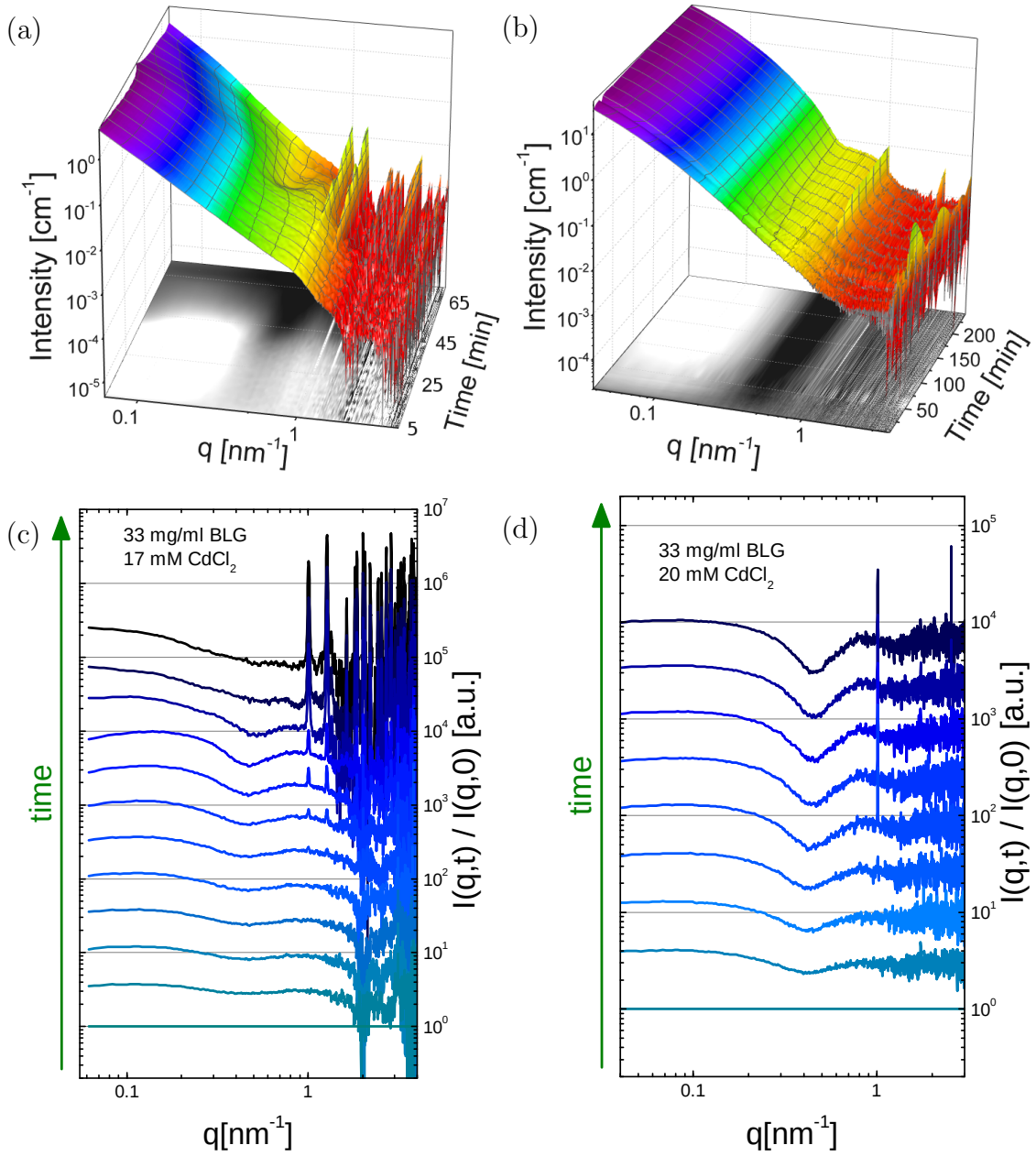

Figure 5.7: *Protein crystallization followed by real-time SAXS measurements. Samples contain 33 mg/mL BLG with 17 mM $CdCl_2$ (a&c) and 20 mM $CdCl_2$ (b&d), located in the transition zone of pseudo-c** (i.e. at the upper end of regime II). (a&b) Three-dimensional plot of $I(q,t)$. The bottom projections are created from $I(q,t)/I(q,t=0)$ and thus monitor the evolution of the curves with time. (c&d) Selected $I(q,t)/I(q,t=0)$ curves for the evolution of Bragg peaks and the amorphous peak. Data are shifted upward by a constant factor for better visibility.*

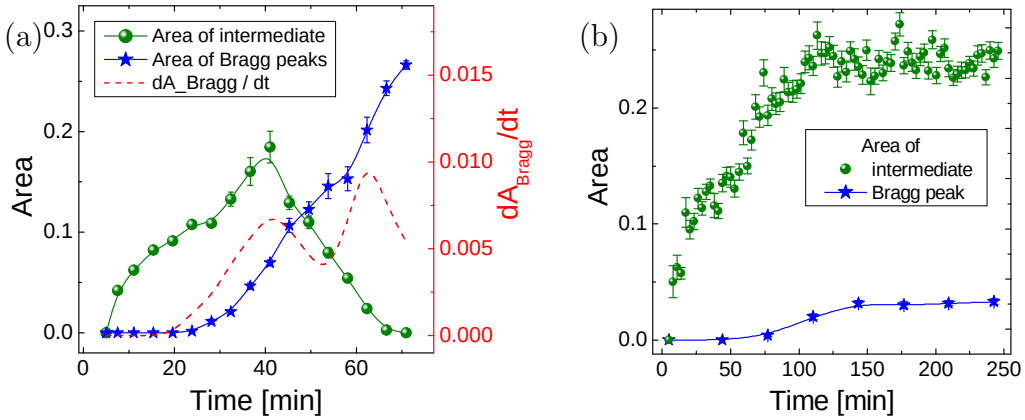

Figure 5.8: *Time-dependence of the area of the broad peak (A_{interm}) and the Bragg peaks (A_{Bragg}). Green spheres (left axis) show $A_{interm}(t)$, the integral of the broad Gaussian function connected to the intermediate. Blue stars (left axis) show $A_{Bragg}(t)$, the integral of the two Bragg peaks in this area. (a) BLG 33 mg/mL with 17 mM $CdCl_2$. For clarity and better statistics, always three data points of $A_{interm}(t)$ and $A_{Bragg}(t)$ were merged into one. The dashed red line in corresponds to the first derivative of the area of Bragg peaks as a function of time (right axis). (b) BLG 33 mg/mL with 20 mM $CdCl_2$.*

followed by optical microscope. However, as seen from Figure 5.3, the number of crystals decreases and the size of crystals increases with increasing salt concentration. This makes the real-time SAXS measurements challenging as the number of the crystals within the illuminated volume drops significantly. We have tried to compensate this by measuring more positions from the sample. This is partly successful, but the time resolution is reduced as only one or two out of ten spots show the development of the Bragg peaks. As shown in Figure 5.8b, less data for Bragg peaks than the MIP are shown. Nevertheless, one can still recognize the interesting kinetics: first the MIP develops relatively fast and becomes saturated after 120 min. Within the current experimental time scale, only a minor fraction of crystalline phase was detected. The experimental observations on the kinetics including particularly the non-monotonous crystallization rate (red dashed line in Figure 5.8a) agree well with a simple model, which will be discussed in the following section.

5.4 Modeling with Rate Equations

In order to compare the observed kinetic features also more quantitatively to possible crystallization scenarios, we employed rate equation models. As essential variables of the modeling, L, I, C_I and C_L denote the mass fractions of free monomers in the liquid, intermediate, crystals in the intermediate and crystals in the liquid. The three paradigmatic cases for the crystallization process shown in

Figure 5.1 are presented as follows:

Classical Nucleation $L \rightarrow C_L$: The classical nucleation theory contains a one-step nucleation process of the critical nucleus from the homogeneous solution. After nucleation, crystallites grow larger from the solution. In terms of modeling, $\Delta_{nL} = k_n L$ and $\Delta_{gL} = k_{gL} L C_L$ represent the nucleation and the growth term with rates k_n and k_{gL}. Using these, the process can be easily modeled by the following set of rate equations:

$$\begin{aligned} \partial_t L &= -\Delta_{nL} - \Delta_{gL} \\ \partial_t C_L &= \Delta_{nL} + \Delta_{gL} \end{aligned} \tag{5.1}$$

Parallel Process $I \leftrightarrow L \rightarrow C_L$: In this nonclassical process, we assume that in addition to the one-step crystallization process, a reversible intermediate is formed in the solution, competing with crystallization for the free monomers. In addition to the nucleation and growth term from the classical one-step nucleation, we include the formation of the intermediate with the term $\Delta_I = k_I(L - I)$. The corresponding set of rate equations reads

$$\begin{aligned} \partial_t L &= -\Delta_I - \Delta_{nL} - \Delta_{gL} \\ \partial_t I &= \Delta_I \\ \partial_t C_L &= \Delta_{nL} + \Delta_{gL} \end{aligned} \tag{5.2}$$

Two-step Nucleation Process $L \rightarrow I \rightarrow C_I \rightarrow C_L$: For the true two-step nucleation, the formation of an intermediate is followed by the nucleation of crystallites within the intermediate. After the subsequent growth of the crystallites has consumed the intermediate, crystals emerge into the remaining liquid and grow further. Table 5.1 describes the steps in more detail and lists the rate terms. Using these, the rate equations read

$$\begin{aligned} \partial_t L &= -\Delta_I - \Delta_{gL} \\ \partial_t I &= \Delta_I - \Delta_{nI} - \Delta_{gI} \\ \partial_t C_I &= \Delta_{nI} - \Delta_e + \Delta_{gI} \\ \partial_t C_L &= \Delta_e + \Delta_{gL} \end{aligned} \tag{5.3}$$

The results of the three models are shown in Figure 5.9 with the model parameters from another study[160] (see also figure caption). For the classical nucleation (Figure 5.9a), nucleation and growth of crystals speed up, until the liquid is halfway consumed, and saturate until complete consumption of the liquid. In the parallel process (Figure 5.9b), the intermediate competes with crystallization and slows down the crystal nucleation and growth considerably. It is important to notice, however, that the crystallization rate is homogeneously increasing until saturation. By contrast, the two-step process (Figure 5.9c) shows the formation of

	Processes	Rates	Parameters
	1. Formation of intermediate until the remaining solution is stable	$\Delta_I = k_I(L - L_0)_+$	L_0 mass fraction of stable solution k_I formation rate
	2. Crystal nucleation within the intermediate	$\Delta_{nI} = k_n I$	k_n nucleation rate
	3. Crystal growth within the intermediate	$\Delta_{gI} = k_{gI} I C_I$	k_{gI} growth rate in intermediate
	4. Crystals emerge into the liquid once intermediate is consumed	$\Delta_e = k_e C_I(\alpha_I C_I - I)_+$	α_I critical ratio of intermediate and crystals k_e emergence rate
	5. Crystal growth in the liquid	$\Delta_{gL} = k_{gL} L C_L$	k_{gL} growth rate in liquid

Table 5.1: *Summary of the proposed two-step process. From the intermediate formed firstly from the liquid, crystals nucleate and grow until they consume the intermediate. Subsequently, the crystals emerge into the liquid and grow from the free monomers.* $(x)_+$ *stands for* x *for positive* x *and zero otherwise.* L, I, C_I *and* C_L *denote the mass fractions of free monomers in the liquid, intermediate, crystals in the intermediate and crystals in the liquid, respectively.*

a plateau in the crystal mass fraction. After increasing at small times, the crystallization rate drops again considerably, whereas it increases monotonically for both one-step cases until saturation (Figure 5.9d). This particular feature of the two-step process is caused by the nucleation and slow growth in the intermediate, while the crystals grow faster once emerged into solution. Thus, the occurrence of the plateau indicates the presence of a multi-step nucleation process.

The large number of model parameters does not allow for a reliable extraction of nucleation rates via model fits to the kinetic analysis of the SAXS data. We emphasize that this is not a problem of the model or the data quality, but the complex pathway that involves many coupled processes. As an example, the amount of crystal nucleation within the MIP depends not only on the rate k_n, but also on the amount of MIP. The latter is mainly determined by the supersaturation represented by L_0, at least for fast formation of the MIP compared to crystal nucleation. Only based on fitting the kinetic data from SAXS, an decrease in L_0 will sequentially cause a decrease in k_n. While thus additional information e.g. on the supersaturation are required for a real quantitative fit, the qualitative signatures of the two-step process such as the plateau in the crystal fraction can still be used to provide evidence for the scenario of a two-step nucleation process.

Nevertheless, this simple model can reproduce the experimental crystallization kinetics at different conditions. This can be achieved either by varying the rate

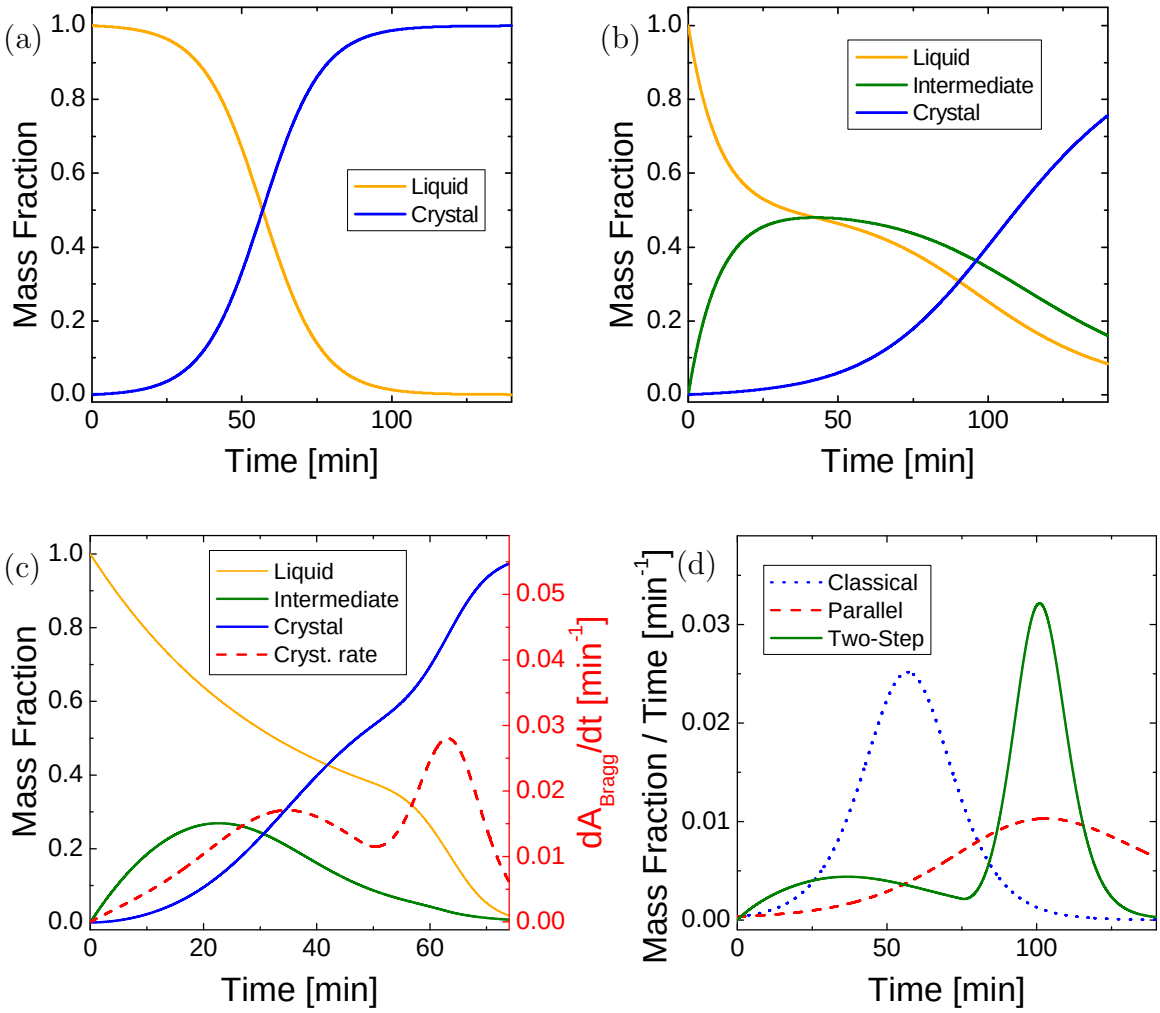

Figure 5.9: *Comparison between different crystallization scenarios: (a) Classical nucleation from a homogeneous liquid and subsequent growth of the crystallites. (b) Parallel processes, representing the formation of an intermediate and the independent nucleation from the liquid with subsequent growth of the crystallites. (c) Two-step nucleation process involving the formation of an intermediate from which the crystals nucleate (see Table 5.1 for further details). The following parameter values were used to show the good qualitative agreement of the model with the data set in Figure 5.8a: $k_e = 0.15\,min^{-1}$, $L_0 = 0.2$, $\alpha_I = 0.2$, $k_{gL} = 0.6\,min^{-1}$, $k_I = 0.03\,min^{-1}$, $k_n = 0.02\,min^{-1}$, $k_{gI} = 0.2\,min^{-1}$. An additional model plot reproducing the data set in Figure 5.8b can be found in the S.I., Figure C.3). (d) Comparison of the crystallization rates dC/dt. While one-step and parallel nucleation processes show a monotonous speed-up until saturation, the two-step process can have a non-monotonous signature with two maxima. The model parameters were chosen for all three models to be the same: $k_I = 0.05\,min^{-1}$, $k_n = 0.02\,min^{-1}$, $k_{gI} = 0.1\,min^{-1}$, $k_{gL} = 0.2\,min^{-1}$, $k_e = 1.0\,min^{-1}$, $L_0 = 0.7$, $\alpha_I = 0.2$.*

parameters or by choosing the amount of MIP which approximately approaches $L - L_0 = 0.2$. This value can be determined experimentally by following the protein concentration in the supernatant with time using UV-visible spectroscopy. A tentative experiment for sample with 20 mM $CdCl_2$ leads to the value of 0.2. Further experiments are needed to refine this parameter.

5.5 Conclusions and Outlook

We have investigated the two-step nucleation process of protein crystallization in solutions by following the overall crystallization kinetics using real-time optical microscopy and SAXS. The experimental results together with a rate equation model provide evidence of two-step nucleation in the early stage of crystallization. The BLG-salt ($CdCl_2$) solutions were chosen at the transition zone of $pseudo-c^{**}$, where small aggregates form after sample preparation. These protein aggregates serve as the metastable intermediate phase (MIP) during crystallization.

SAXS and SANS reveal that the MIP shows a certain local ordering instead of random aggregates as monitored by a broad shoulder at intermediate $q \approx 0.7\,\mathrm{nm}^{-1}$, and a monomer-monomer correlation peak at q around $2\,\mathrm{nm}^{-1}$. Real-time SAXS results show that the crystallization kinetics is proportional to the development of the MIP in the early stage of crystallization, i.e. the appearance of a local maximum in the crystallization rate at the maximum quantity of the intermediate. In the late stage of crystallization, a plateau is developed due to the transition from nucleation controlled in the early stage to growth controlled after the consumption of MIP. This transition in the overall crystallization kinetics is a typical feature for the two-step nucleation in the early stage. These experimentally observed kinetics can be reproduced using a rate equation model.

For further real-time measurements, we note that the smaller beam size and scattering volume of SAXS can be compensated by using SANS. The combination of real-time SAXS and SANS can provide more systematic information of the crystallization kinetics.

A. Sauter acknowledges a fellowship of the Landesgraduiertenförderung. The authors thank T. Stehle (IFIB, Universität Tübingen) for sharing lab resources. This work was supported by the DFG.

Chapter 6

Publication D. *Structural Evolution of Metastable Protein Aggregates in the Presence of Trivalent Salt Studied by (V)SANS and SAXS*

(submitted to Journal of Applied Crystallography)

Andrea Sauter, Fajun Zhang, Noemi K. Szekely, Vitaliy Pipich, Michael Sztucki and Frank Schreiber

Contributions	
Research design	*AS, FZ, FS*
Experiments	*AS, FZ*
Technical Assistance	*NKS, VP, MS*
Data analysis and interpretation	*AS, FZ*
Paper writing	*AS, FS, FZ*

We present a study of the structural evolution of protein aggregates in solutions of a globular protein, β-lactoglobulin (BLG), in the presence of YCl_3. These aggregates are often observed before crystallization starts and metastable with respect to the crystalline phase. Here we focus on the characterization of the structures of this intermediate phase at different length scales using a combination of VSANS, SANS and SAXS. Our results reveal a hierarchical structure from monomer, dimer, compact cluster to fractal protein aggregates. Upon cooling, the overall hierarchical structure is preserved, however, the evolution of the internal structure within the aggregates with temperature is clearly visible: the monomer–monomer correlation peak disappears completely, the cluster–cluster correlation is enhanced. At a larger length scale, the fractal dimension of protein aggregates increases. Moreover, the time-dependent SAXS measurements during a temperature ramp show clear isosbestic points and the kinetics of the structural evolution of the metastable intermediate phase can be well described using a two-state model.

6.1 Introduction

Protein crystallization, as an important research field, provides not only the most successful way for determining high resolution 3D structures of bio-macromolecules, but also rich conditions for understanding the early stage of crystallization. For example, it is very common to observe oil-like droplets in the crystallization assay before crystals appear. This liquid–liquid phase separation (LLPS) in protein solutions has been demonstrated to be metastable with respect to the crystalline phase, thus the co-existence line is located below the solubility line in a typical phase diagram of protein solutions[26,38,54,126,212]. In addition, mesoscopic protein clusters have been observed in concentrated solutions for several proteins[64,96,137,181]. Counter-balanced interactions have been reported to lead to equilibrium clusters as well as dynamic clusters in concentrated protein solutions[101,177,184]. Importantly, the existence of protein clusters or a metastable LLPS may change the kinetic pathway of crystal nucleation dramatically.

The pioneering work by ten Wolde and Frenkel showed that far from the metastable LLPS region nucleation follows the classical mechanism, i.e. nuclei form directly in a supersaturated solution with the structure and density of the final crystalline phase[7,56,172]. However, when approaching the critical point, dense liquid-like droplets form first which further follow a structural change towards crystallization[196]. In the last decade, increasing evidence has shown that clusters, amorphous nanoparticles and other precursors may serve as an important intermediate for crystallization[13,50,64,65,122,131]. Theoretical studies and simulations suggest that the existence of a metastable intermediate phase can promote crystal nucleation[79,196,198]. Huge efforts have been focused on characterizing the structure of these intermediate phases for a better understanding of the early stage of crystallization and the role of the intermediate phase on crystal nucleation[30,64,65,100,137,138,161,174,181,228].

In our previous work, we found that multivalent metal ions provide an efficient way in tuning the effective protein–protein interactions in solution, leading to complex phase behavior including reentrant condensation (RC), metastable LLPS, cluster formation and crystallization[151,183,220,222,224,226,227]. The rich phase behavior induced by these metal ions provides not only the optimized conditions for protein crystallization, but also model systems for studying the *pathways* of protein crystallization[158–160]. Figure 6.1 shows the experimental phase diagram of β-lactoglobulin (BLG) in the presence of YCl_3. We observe the so-called reentrant condensation behavior, i.e. a phase separated state (Regime II) exists in between two boundary salt concentrations $c^* < c^{**}$, outside of regime II, the samples are clear. The physical mechanism of this RC behavior is due to the effective charge inversion of proteins and a cation-mediated attraction[152,227]. Similar phase behavior has been observed for several proteins in solution in the presence of trivalent metal ions[151,220,224,226].

Studies on the protein crystallization in different regions of Figure 6.1 have shown that close to c^*, crystallization takes place without visible precursor. Close

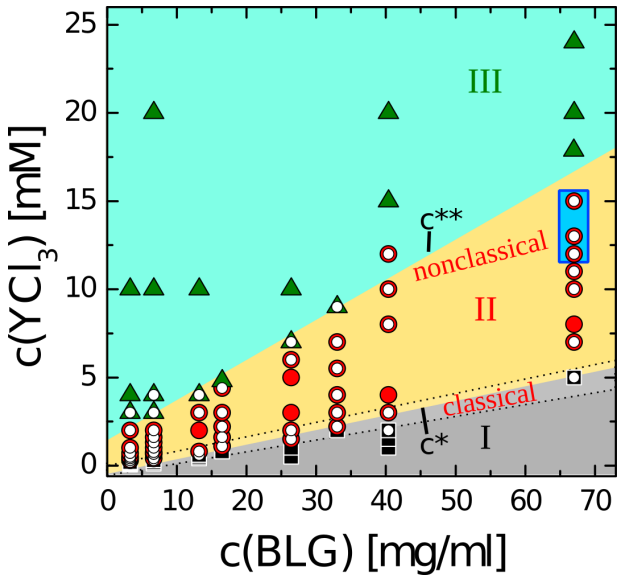

Figure 6.1: *Experimental phase diagram of BLG in the presence of YCl_3 determined at 21°C. Regimes I, II, III corresponds to clear, turbid and clear state, respectively. Conditions in regime I are plotted as black squares, conditions in regime II as red circles and conditions in regime III as green triangles. Small white circles mark samples that crystallized after one week at 4°C. The blue rectangular marks sample conditions we focus on in this work. Different crystallization pathways are observed. Close to c*, crystallization takes place without visible precursor, which is labeled as classical; Close to c**, aggregates or a dense liquid phase are observed before crystallization, which is labeled as nonclassical.*

to c^{**}, samples have a transition temperature, T_{tr}. When samples are prepared at temperatures above T_{tr}, the solution is clear. Below T_{tr}, the solution becomes turbid immediately by forming large protein aggregates or a dense liquid phase, depending on the sample conditions[158,221,227]. After an induction time, nucleation starts and crystal growth consumes the intermediate phase. In the end, only crystals are left[158,227].

One of the interesting observations on the intermediate states is the morphology change of protein aggregates with time. Optical microscopy observations clearly show that the turbid solutions with amorphous protein aggregates develop into a network of aggregates which further relaxes into a liquid-like structure[158]. Based on the fact that these areas are growing and merging, they are most likely in a liquid state instead of a gel[158,227], although this is still an open question. In fact, little is known about the internal structures of the intermediate and their development with time. This information may be crucial for the subsequent nucleation. For a deeper understanding of the role of the metastable intermediate phases in protein nucleation and crystallization, we performed a systematic study of the structures and their evolution with time and temperature by small angle X-ray and neutron scattering. In particular, we aim to characterize the structural fea-

tures of these protein aggregates or liquid-like networks at different length scales and the kinetics of the structural evolution.

6.2 Experimental Section

6.2.1 Materials

The globular protein β-lactoglobulin (BLG) from bovine milk (*product no. L*3908) was purchased from Sigma-Aldrich. This product is a mixture of the genetic variants A and B, that differ at two positions in the primary sequence of 162 amino acids in total[207]. At room temperature and neutral pH, BLG is acidic (isoelectric point of 5.2) and predominantly in dimer state[44,207]. YCl_3 was also purchased from Sigma-Aldrich as a anhydrous powder with a purity of 99.99% (*product no.* 451363).

For sample preparation, appropriate amounts of salt stock solution, degassed MilliQ (18.2 MΩ, Merck Millipore) water and protein stock solution were mixed. Stock solutions were prepared by dissolving the salt or protein powder in deionized (18.2 MΩ) and degassed MilliQ water. The protein concentration of stock solutions was determined by UV absorption measurements using an extinction coefficient of $0.96\,\mathrm{l}\cdot\mathrm{g}^{-1}\cdot\mathrm{cm}^{-1}$ at a wavelength of 278 nm[182]. All samples in this work were prepared without additional buffer since buffers can affect the phase behavior of proteins and the solubility of salts. The pH of the solutions was monitored using a Seven Easy pH instrument from Mettler Toledo. The pH values for all experimental conditions were above the $pI = 5.2$ of BLG[44].

6.2.2 Methods

Fourier transform infrared spectroscopy (FTIR) (IFS 48 from Bruker) was applied to monitor the stability of the protein secondary structure under the experimental conditions as a function of temperature, salt concentration and further treatment. Samples in the same composition but in heavy water (D_2O) were prepared for the better signal-to-noise ratio.

Small angle X-ray scattering (SAXS) experiments were performed on beamline ID2 at the ESRF (Grenoble, France) using a sample-to-detector distance of 2 m. The beam energy was set to 16 keV, the accessed q values ranged from 0.006 to $0.34\,\mathrm{Å}^{-1}$. The data sets were reduced by subtracting the scattering of pure salt solution as a background and normalized to absolute intensities. The samples were measured using a quartz cell with a diameter of 2 mm. For the time-resolved measurements, the exposure time was 0.1 s, the waiting time between measurements was 300 s, and the data saving and transfer took about 6 s. More detailed information on q-resolution calibration and data reduction can be found in Ref. [80,129].

Small angle neutron scattering (SANS) was performed at FRM2 (Garching, Germany) at the instrument KWS-2. The utilized neutron wavelength was 4.5 Å

with $\Delta\lambda/\lambda$=20%. The sample-to-detector distances were 2, 8 and 20 m. The resulting total q-range was $2.2{\cdot}10^{-3}$ to 0.34 $Å^{-1}$.

Very small angle neutron scattering (VSANS) was performed at FRM2 (Garching, Germany) at the instrument KWS-3 with a sample-to-detector distance of 10 m and allowed us to cover a large q range. The applied neutron wavelength was 12.8 Å with $\Delta\lambda/\lambda$=20%, resulting in a q-range from $2.3{\cdot}10^{-4}$ to $2.1{\cdot}10^{-3}$ $Å^{-1}$.

Samples for neutron scattering were prepared in D_2O for better contrast. Intensity profiles were obtained by correcting the 2D intensity pattern to absolute scale and azimuthal averaging. The solvent background was subtracted afterwards. Data reduction and absolute intensity calibration was performed using the program QtiKWS provided by JCNS[144]. Data fitting was preformed using OriginPro (two-state analysis) and IGOR Pro with macros provided by NIST[90].

6.2.3 Data Analysis

The scattering intensity of SANS experiments of protein solutions can be described as

$$I(q) = N_p(\Delta\rho)^2 V_p^2 P(q) S(q) + B_{inc} \tag{6.1}$$

where N_p is the number density of proteins per unit volume; V_p is the volume of a single protein; $\Delta\rho$ is the difference in scattering length density between protein and the solvent; B_{inc} is the incoherent background; $P(q)$ and $S(q)$ are the form factor and the structure factor, respectively. The momentum transfer is defined as $q = 4\pi/\lambda \cdot sin(\theta)$ with the scattering angle 2θ. The size and shape of structures at different length scales, such as the native dimer and aggregates, are described using a spherical or ellipsoidal form factor. The effective interactions are described using structure factors calculated from different potentials that are listed in the following.

The hard sphere potential is defined by

$$U_{HS}(r) = \begin{cases} \infty & r \leq \sigma \\ 0 & r > \sigma \end{cases} \tag{6.2}$$

with the hard sphere diameter of σ. The corresponding structure factor is obtained by solving the Ornstein–Zernike (OZ) equation numerically with the Percus–Yevick (PY) closure[9].

The sticky hard sphere potential is used to describe the short range attraction between protein clusters, which is defined by[15]

$$U_{SHS}(r) = \begin{cases} \infty & r < \sigma \\ -u & \sigma < r < \sigma + \Delta \\ 0 & r > \sigma + \Delta \end{cases} \tag{6.3}$$

where τ is the stickiness parameter

$$\tau = \frac{1}{12\epsilon_\Delta} exp(-u/kT) \tag{6.4}$$

with the perturbation parameter ϵ_Δ,

$$\epsilon_\Delta = \Delta/(\sigma + \Delta). \tag{6.5}$$

Δ is the width of the square well, and u is the depth[90].

The screened Coulomb potential is used to describe the long range electrostatic repulsion of protein clusters due to the accumulated charges, which is defined by[73,77]

$$U_{SC}(r) = \begin{cases} \frac{Z^2 e^2}{\epsilon(1+\kappa_D \sigma/2)^2} \frac{exp(-\kappa_D(r-\sigma))}{r} & r > \sigma \\ \infty & r \leq \sigma \end{cases} \tag{6.6}$$

with Z the protein surface charge, e the electronic charge, and ϵ the dielectric constant of the solvent. κ_D is the inverse of the Debye screening length, determined by the ionic strength of the solution and σ the effective diameter of the ellipsoid with the axes a, b, b.

The Beaucage model was applied which is well suited for the analysis of small-angle scattering data from complex systems that contain multiple levels of related structural features like fractal and particulate systems. It yields multiple sets of a radius of gyration and a power law from different structural levels by modeling the Guinier and Porod regions[16,17,72]. The multiple levels of scattering intensity can be expressed as

$$\begin{aligned} I(q) \approx \sum_{i=1}^{n} \Bigg\{ & G_i \cdot exp\left(\frac{-q^2 R_{g(i)}^2}{3}\right) \\ & + B_i \cdot exp\left(\frac{-q^2 R_{g(i+1)}^2}{3}\right) \cdot \left[erf\left(\frac{q R_{g(i)}}{\sqrt{6}}\right)^3 \cdot \frac{1}{q} \right]^{d_i} \Bigg\}. \end{aligned} \tag{6.7}$$

It is characterized by three principal fitting parameters per level: a radius of gyration R_g, a Porod exponent d and a Guinier scaling factor G (from which a Porod scaling factor B can be derived)[72].

6.3 Results

6.3.1 Hierarchical Structure Characterized by (V)SANS and SAXS

We first present an overview of interactions and structures in protein solutions as a function of salt concentration and temperature. In order to study the struc-

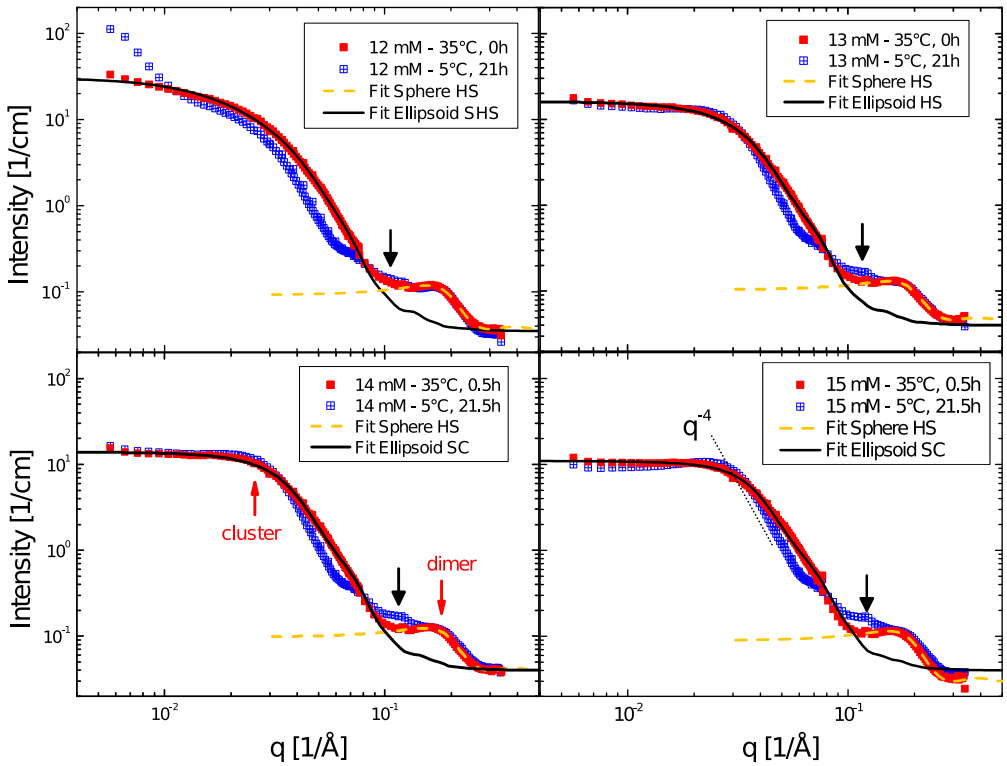

Figure 6.2: *Structure and interactions of protein-salt mixtures characterized by SANS in dependence of temperature and salt concentration. The protein concentration was fixed to 67 mg/mL, the salt concentrations were 12, 13, 14 and 15 mM YCl_3. Data collected at two temperatures, 35 °C and 5 °C, are presented. The times given in the legend are the starting times. The measurements themselves lasted 2.5 hours each. The black arrows label an additional structural change that is more pronounced with high salt concentration. Model descriptions for different q ranges are shown in lines and discussed in detail in the main text.*

tural evolution as a function of temperature (observable range 0-40 °C), we chose samples with transition temperatures T_{tr} close to or below 35 °C, samples with a BLG concentration of 67 mg/mL and YCl_3 concentrations between 12 and 15 mM. According to the experimental phase diagram in Figure 6.1 which was obtained at room temperature, these samples are in the upper area of regime II. Increasing temperature shifts the samples into Regime III. When prepared at 35 °C, the samples with 13-15 mM salt were clear, only the sample with 12 mM salt was turbid indicating a higher T_{tr}. When cooled down to 5-10 °C, all samples became turbid.

Figure 6.2 shows SANS results for these samples at 35 °C and 5 °C. Measurements between the first and the last run are shown in the supporting material (Figure D.1). At 35 °C, the SANS curves for all samples are similar in the high q region, but the intensity at the low q region decreases with increasing salt concentration. For the sample with 12 mM salt, the low q upturn is consistent with the observation that this sample is already turbid at 35 °C, which becomes more

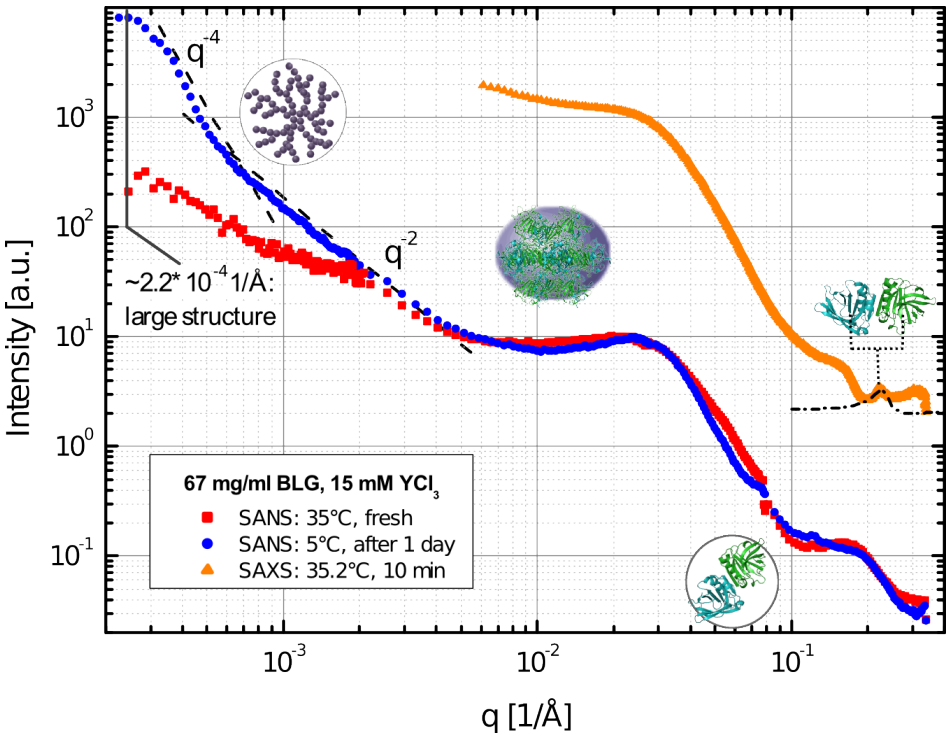

Figure 6.3: *Hierarchical structure of protein aggregates characterized by (V)SANS and SAXS for samples with 67 mg/mL BLG and 15 mM YCl_3 at 35 and 5 °C. The merged SANS data are from KWS-3 (VSANS) and KWS-2 (SANS). SAXS data of the same sample at 35 °C show an additional monomer-monomer correlation in the high q region. The structural interpretation at different length scales is also shown.*

significant at 5 °C due to the formation of more compact large aggregates. The other samples, however, do not have such strong intensity increase at the low q region of this q range even after 21h at 5 °C.

The SANS measurement curves at 35 °C show two main steps in intensity at 0.03 and 0.18 $Å^{-1}$, which can be explained by the contribution of small protein clusters and the native dimer of BLG, respectively, consistent with previous measurements on this system[221]. The dimer contribution has developed a visible dimer–dimer correlation peak. It was further analyzed using a simple spherical form factor and a hard-sphere potential. The radius of the spherical dimer form factor is about 15 Å and was fixed for further data fitting. The volume fraction is between 0.22 and 0.23 and slightly decreases with increasing salt concentrations. The detailed fitting results are shown in the supporting material (Table D.1).

For the cluster contribution, the scattering intensity in the low q region (below 0.03 $Å^{-1}$) decreases with increasing salt concentration. A correlation peak appears which becomes even stronger at lower temperature. With 14 and 15 mM YCl_3, the shoulder at 0.03 $Å^{-1}$ develops to a cluster–cluster correlation peak. At q values between 0.04 and 0.07 $Å^{-1}$, all curves show an asymptotic linear trend with slopes

between -3.5 (start) and -4 (end). This behavior satisfies Porod's law and implies that the scatterers at this length scale (clusters) have a compact, non-fractal structure with a sharp interface. A model fit of the curves for 14 and 15 mM YCl_3 using a screened Coulomb potential together with an ellipsoid form factor shows that the overall net charge increases with increasing salt concentration. While the charges are nearly constant with decreasing temperature, the effective size of the ellipsoid form factor and the volume fraction increase (Table D.1). With 13 and 12 mM YCl_3, the data is better described using the hard-sphere and sticky hard-sphere potential, respectively, indicating that the effective interactions between the clusters undergo a transition from repulsion to attraction with lowering salt concentration.

The development of the cluster-cluster correlation can be understood as the accumulation of net charge. Studies on the reentrant condensation have demonstrated the effective charge inversion of proteins upon increasing salt concentration crossing the second regime[220,226]. When these samples are prepared at high temperature (35 °C), they are in the third regime, i.e. the effective surface charge is already positive, but the strong bridging effect of the metal ions leads to still attractive overall interactions. However, with increasing the size of clusters, the net charge of the clusters accumulates, and when the charge is high enough, a further increase of the cluster size will be prohibited by the re-established Coulombic repulsion.

While there is no visible change for the dimer contribution during cooling, the slopes for the clusters decrease, leading to a new shoulder visible at 0.08 $Å^{-1}$ at the low temperature for all samples. Additional structure change at 0.12 $Å^{-1}$ (labeled using a black arrow) occurs at lower temperature and is more pronounced with high salt concentration. These new features may correspond to structures of $D = 2\pi/q \approx 78$ and 52 Å, respectively. The formation of this hierarchic structures also suggests that proteins within the clusters have a certain flexibility for rearrangement upon cooling.

The hierarchical structure of the protein aggregates for samples with 67 mg/mL BLG with 15 mM YCl_3 were further characterized using VSANS and SAXS to cover a broader q-range. As presented in Figure 6.3, the combined (V)SANS and SAXS data at 35 and 5 °C provide the structure information from monomer–monomer correlation, compact clusters up to fractal aggregates in length scales from nanometer to µm. These structures at different length scales can be described with the assembled structures of protein (schematics) in the figure.

The extended VSANS data in the low q region clearly show the increase in intensity at both temperatures. The asymptotically linear increase in intensity at lower q suggests larger protein aggregates that are more fractal than the clusters at smaller length scale. The scattering intensity from (mass) fractal aggregates is given by[140] $S(q) \propto q^{-d}$. The fractal dimension, d, determined from the slopes are around 1.7 and 1.9 before and after cooling in the q range of (0.5 - 4)·10^{-3} $Å^{-1}$. At even lower q, after cooling, another Porod region becomes visible and a plateau at around 2·10^{-4} Å is visible, which indicates that the fractal aggregates can further

organize in an even larger and more compact structure with a mean size in real space of $D = 2\pi/q \approx 3\,\mu m$.

SAXS data show additional structural features due to the high resolution in the high q region. A peak at $0.22\,Å^{-1}$ is clearly visible in the SAXS curve (not visible in SANS) after preparation at 35 °C, which can be attributed to the monomer–monomer correlation within clusters as discussed in our previous work [221]. The monomer–monomer correlation in the initial state is fitted using a spherical form factor with a radius of 15 Å and a hard sphere structure factor, which gives a volume fraction of 0.50. The peak completely disappears during the cooling process. Compared to the SAXS data, the SANS data are shifted towards higher q, mainly due to the contribution of the hydration shell, to which only SAXS is sensitive [221]. SAXS measurements of the whey protein microgel obtained by thermal treatment also reveal a hierarchical and fractal structures [123,166]. Interestingly, it also shows a structural feature at $\sim 0.07\,Å^{-1}$, which has been explained by the close neighbor effect of the denatured whey protein monomers within the microgel.

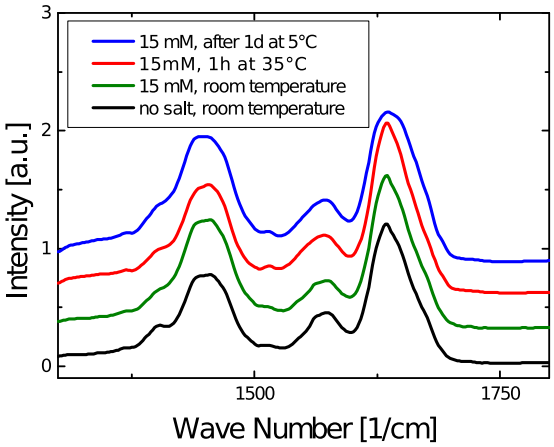

Figure 6.4: *FTIR spectra of protein samples with a BLG concentration of 67 mg/mL: (Black) without salt and prepared at room temperature; (Green) with 15 mM YCl_3 and prepared at room temperature; (Red) with 15 mM YCl_3 and prepared and equilibrated at 35-38 °C for 1 h; (Blue) with 15 mM YCl_3, prepared at 35 °C and quenched and stored at 5 °C for a day.*

We emphasize that the observed protein condensation is not caused by a change of the protein structure induced by YCl_3 and further treatments. Figure 6.4 shows FTIR spectra for a sample with a BLG concentration of 67 mg/mL and 15 mM YCl_3 at room temperature and at 35-38 °C (measurements for 12-14 mM can be found in the supporting information, Figure D.2). For comparison, the spectrum of a protein solution with the same protein concentration but without added salt is also shown. As one can see, in all cases, the amide I and amide II bands are in the same shape and ratio of intensity. These measurements confirm that there is no significant change on the secondary structure of protein neither caused by addition of YCl_3 nor by temperature change. Moreover, the successful growth of

high-quality crystals and fine structural analysis confirm that the proteins are still in their native state[158,227].

6.3.2 Structural Evolution Followed by Time-Resolved SAXS

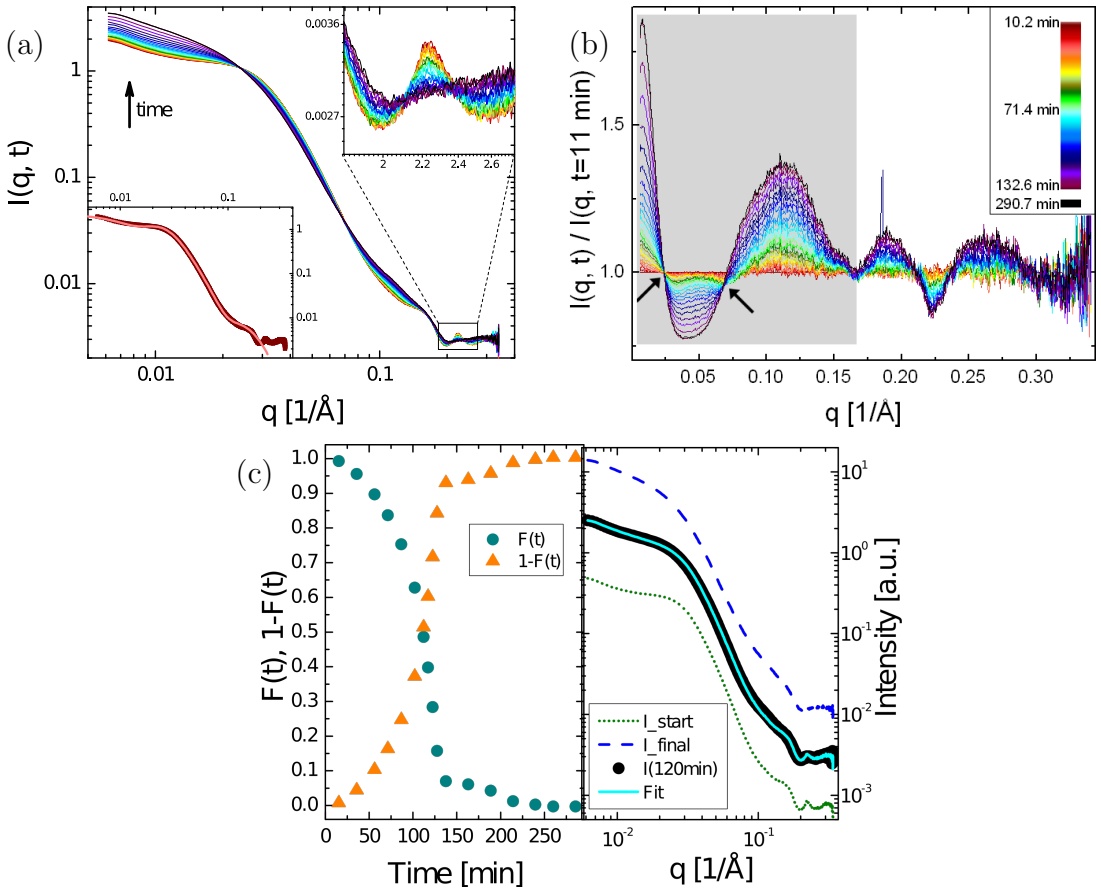

Figure 6.5: *(a) Real-time SAXS profiles of 67 mg/mL BLG with 15 mM YCl_3 at different times during a temperature ramp from 37 to 10 °C. Insets show the zoom-in of the monomer–monomer correlation peak (up-right) and a typical SAXS data curve with a two-level Beaucage fit (bottom-left). (b) Modified SAXS curves from a) by dividing by the first curve. (c) Two-state analysis of the structural evolution kinetics using Equation 6.8. Left: plots of $F(t)$ and $1-F(t)$ as a function of time. Right: example SAXS curve 102 min after the preparation (black) with obtained fit (cyan) along with the initial and final curve (shifted up and down for better visibility).*

Optical microscope observations indicate that the protein aggregates formed after sample preparation can further relax into a liquid-like network within hours[158]. Later crystallization occurs mainly at its interface. The hierarchical structure at

different temperature presented in the previous section indicates the local structure change with time and temperature. In the following, the real-time SAXS data are further analyzed in order to understand the transition kinetics.

The real-time SAXS data including 58 scans in steps of 5 min during a temperature ramp from 37 to 10 °C are presented in Figure 6.5a. In general, during the temperature ramp, the overall scattering profile is preserved, which is consistent with the SANS measurements in Figure 6.2. However, local structure changes are clearly visible. For example, the intensity for q below 0.025 Å^{-1} increases with time (or decreasing temperature), whereas the intensity decreases with time in the q range between 0.025 Å^{-1} and 0.075 Å^{-1}, resulting an isosbestic point. Furthermore, the monomer–monomer correlation peak reduces with time and finally disappears. These changes are better presented in Figure 6.5b, where the SAXS data were divided by the first curve (about 11 min after sample preparation). This figure therefore displays the changes in this process. At 0.025 Å^{-1} and 0.075 Å^{-1}, respectively, the intensity remains constant upon temperature change and time, characterizing isosbestic points. At adjacent lower q, the intensity is decreasing, and increasing on the other side. The existence of an isosbestic point suggests an equilibrium of two populations of aggregates with one growing in number at the cost of the other one[135].

Isosbestic (or iso-scattering) points can occur if a system transforms from one state to the other and the scattering intensity for both states stays constant at (at least) one given q[135]. Therefore, the curve can be described as the sum of two constituents[135]:

$$I(q) = F(t) \cdot I_{start}(q) + [1 - F(t)] \cdot I_{final}(q) \tag{6.8}$$

$$F(t) = (I - I_{final})/(I_{start} - I_{final}), \tag{6.9}$$

A two-state analysis was performed using Equation 6.8 to fit the SAXS curves measured at different times with $I_{start}(q)$ the first curve in time and $I_{final}(q)$ the last one. $F(t)$ describes the variation of the scattering intensity. The obtained $F(t)$ and $1 - F(t)$ are plotted in Figure 6.5c left. Using the respective F, all the intermediate SAXS curves can be modeled with a high precision. The right part of Figure 6.5c shows an example for the curve 102 min after the preparation together with the obtained fit. The coefficient of determination R^2 (0 for no correlation, 1 for perfect correlation) was between 0.99977 and 0.99999 for different curves. This analysis provides further evidence that the internal structural evolution within the metastable protein aggregates is formed via a simple two-state process. It is worth noting that in our previous work[221], a similar sample has been measured using SAXS and Bragg peaks were observed after two cycles of a temperature ramp from 25 to 10 °C. The SAXS profiles in the present study are very similar, but no crystals are observed after the measurements. The exact reasons are unknown, but the different thermal history may be one of them.

As we have pointed out, the temperature dependent monomer–monomer cor-

relation peak may be related to the flexibility of proteins within clusters and aggregates. We have further analyzed the monomer–monomer correlation peak by fitting it using a Gaussian function (after subtraction of a linear background). The Gaussian was integrated in q and the area was plotted as a function of time (or temperature, respectively) (see Figure 6.6a). The value of the area (also the shape of the correlation peak) is nearly constant before it starts to decrease at approximately 22 °C. Below ca. 13 °C (after 70 min), this correlation peak disappears completely.

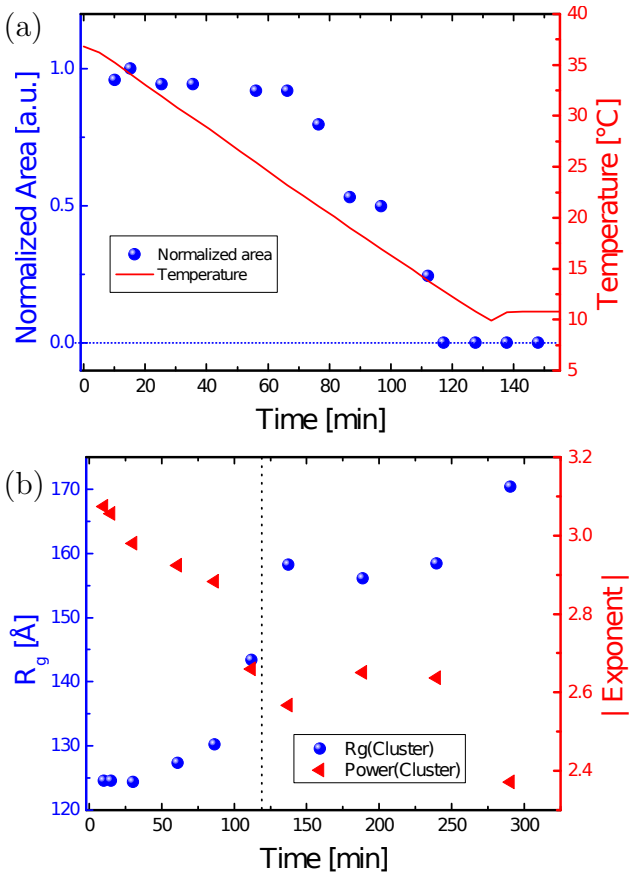

Figure 6.6: *(a) Area of the monomer–monomer correlation peak (blue data points) fitted by a Gaussian. The red solid line shows how the temperature was changed with time. (b) Radius of gyration and absolute value of the exponent for the cluster shoulder of the Beaucage fits. The dashed line marks where the monomer–monomer correlation becomes invisible.*

Further, the Beaucage model[16,17] was also used to fit the real-time SAXS data (bottom-left in Figure 6.6a). The considered curves show two levels of a Guinier and a Porod region corresponding to the dimer and cluster contribution as also shown in the SANS data. Therefore, two radii of gyration and two exponents are obtained for every measurement run. Detailed fitting results are shown in the supporting material (Table D.2). The radius of gyration of a dimer, $R_g(Dimer)$ is

23.5 Å and nearly constant. The radius of gyration of a cluster, $R_g(Cluster)$, grows significantly in size until approximately 175 min (Figure 6.6b, left axis). Then a slower increase is visible. $R_g(Cluster)$ is 125 Å at the beginning and 170 Å at the end. The Porod exponent for the cluster contribution in dependence of time can also be found in Figure 6.6 (right axis). After an initial decrease, the exponent from the scattering of clusters becomes approximately constant. This coincides with the disappearance of the monomer–monomer correlation peak and therefore most likely represents the same structural change in the protein solutions with temperature decrease and time.

6.4 Conclusions

A hierarchy of structures of protein aggregates in aqueous solution of BLG and YCl_3 which may serve as the intermediate for protein crystallization has been characterized using (V)SANS and SAXS. Structural features at different length scales including monomer–monomer correlation, the scattering of dimers, compact clusters and a fractal structure in µm length scales have been identified.

When prepared at 35 °C, the monomer–monomer correlation is visible in the SAXS curves. After cooling to 5-10 °C, the dimer acts as elemental building block and the monomer–monomer correlation peak disappears. It is interesting to see that a cluster–cluster correlation peak becomes visible with increasing salt concentration. These rather compact protein clusters might be formed by the balance between the bridging effect of metal ions and the electrostatic repulsion due to the accumulated net charges. The appearance of the isosbestic points in the real-time SAXS data indicates a two-phase process for the internal structure change within the metastable protein aggregates, which is further demonstrated using a two-state analysis for the SAXS data.

Acknowledgement

We gratefully acknowledge funding from DFG and Landesgraduiertenförderung, as well as allocation of beamtime from ESRF and Forschungszentrum Jülich / FRM2 (Munich). We thank Dr. F. Roosen-Runge and Ph.D. P. Fenter for valuable discussions.

Chapter 7

Complementary Study: On the Universality of Inducing Protein Crystallization by RC

Most of the results presented in this thesis were obtained with BLG in the presence of three different multivalent metal salts (YCl_3, $ZnCl_2$, and $CdCl_2$). Other systems with trivalent salts that show RC behavior were found in our group [80,151,220,222,223,226], whereof HSA in the presence of YCl_3 was observed to crystallize [223]. This chapter addresses the question how universal these observations are and if RC could be used as a method to crystallize a large number of proteins. It also contains results obtained by S. Da Vela and O. Matsarskaia. While the proteins tested in Section 7.1 are commercially available and well-studied, Section 7.2 presents a recently isolated enzyme (see G. Zocher 2012 [229]).

The divalent salts $CdCl_2$ and $ZnCl_2$ can, as shown in the publications A-C, also induce RC and crystallization in BLG, but only a first boundary and no crystallization can be observed in both HSA and BSA. With $CoCl_2{\cdot}6H_2O$ and $SrCl_2$, not even c^* could be induced in BLG and BSA; however, a charge inversion still can be observed. In the following, we focus on trivalent lanthanide salts.

7.1 Toolbox for Tuning Phase Behavior in Protein Solutions Using Trivalent Salts

A systematic study of the effects that rare earth element salts have on several globular, acidic proteins was performed. We especially focused on RC and the ability of the salts to induce crystallization. One of the advantages of rare earth element salts is their highly positive charge (+3), which is an important prerequisite for inducing charge inversion. Monovalent salts do not interact strong enough with proteins [220] and our observations with BLG in the presence of the divalent $CdCl_2$ and $ZnCl_2$ are probably no universal property. Another important advantage of these salts is the fact that they have, in contrast to several other, more common elements such as iron or aluminum, no strong effect on the pH of the solutions due to autoprotolysis and formation of aqua-complexes [115,151]. Gd is additionally interesting due to the paramagnetic properties of Gd(III) complexes [115,150].

7.1.1 Phase Behavior Under Variation of the Cation

The radii of the five ions La^{3+}, Ce^{3+}, Gd^{3+}, Y^{3+}, and Yb^{3+} decrease continuously (see Table 7.1). The larger the ions are, the smaller is their charge density. We therefore assume the strength of the ions in terms of inducing RC to be $Yb^{3+} > Y^{3+} > Gd^{3+} > Ce^{3+} > La^{3+}$. Table 7.1 shows the effect of the trivalent salts $LaCl_3$, $CeCl_3$, $GdCl_3$, YCl_3, and $YbCl_3$ on the negatively charged proteins BLG, HSA, BSA, OVA, and pepsin. Conditions marked by two green checkmarks mean a complete reentrant condensation behavior with a first critical salt concentration c^* and a second critical salt concentration c^{**}. Systems marked by one orange checkmark showed only a first critical salt concentration c^*. A first critical boundary was found in all of the systems tested, a second boundary in most cases. However, pepsin shows no normal c^{**} with all salts tested. Re-clearing was found for some of the salts at very high salt concentrations, for example with 92 mM YCl_3 at a pepsin concentration of 5 mg/mL. For comparison: in other systems with 5 mg/mL protein, c^{**} was usually found between 1.5 and 3 mM salt. We assume that the high salt concentration at which the system re-clarifies has other causes than usual RC (see below where the influence of the pH value is discussed). Only in OVA, a cation-dependent difference is observable: $LaCl_3$, $GdCl_3$ and YCl_3 cause a second boundary c^{**}, while the "strongest" salt judged by the charge density, $YbCl_3$, does not. However, protein material falls out of OVA solutions above c^{**} after some time: these third regimes are not stable.

Table 7.1: *Protein-salt systems tested for RC. Radii are the effective ionic radii from Ref. [176]. Systems with c* and c** are marked by two green checkmarks, systems with only a first boundary by one orange checkmark.* [1]*re-clearing at very high salt concentrations.* [2]*not investigated.*

	$LaCl_3$ La^{3+} (r=103 pm)	**$CeCl_3$** Ce^{3+} (r=101 pm)	**$GdCl_3$** Gd^{3+} (r=94 pm)	**YCl_3** Y^{3+} (r=90 pm)	**$YbCl_3$** Yb^{3+} (r=87 pm)
BLG	✓✓	✓✓	✓✓	✓✓	✓✓
HSA	✓✓	✓✓	✓✓	✓✓	✓✓
BSA	✓✓	✓✓	✓✓	✓✓	✓✓
OVA	✓✓	[2]	✓✓	✓✓	✓
Pepsin	✓[1]	[2]	✓[1]	✓[1]	✓

A linear dependence of c^* on the protein concentration is found (with a certain deviation at low protein concentrations), in accordance with earlier experiments[226]. c^* has been fitted linearly for different protein-salt systems. The slopes of the fits (Table 7.2) increase from $LaCl_3$ to $GdCl_3$ to YCl_3 for pepsin and BLG. However, for HSA and BSA, the inverse behavior is observable. $YbCl_3$, the smallest of the cations tested, deviates from these trends. Other effects such as orbital occupancy might play a decisive role here.

Table 7.2: *Slopes b from linear fits $y = a + b * c(Protein)$ of c^* in various systems. The slope increases from $LaCl_3$ to $GdCl_3$ to YCl_3 (decreasing cation size) for pepsin and BLG. However, for HSA and BSA, the inverse behavior is observable.*

	BLG	Pepsin	HSA	BSA
$LaCl_3$	0.035±0.001	0.041±0.004	0.08±0.01	0.100±0.002
$GdCl_3$	0.063±0.002	0.050±0.001	0.06±0.005	0.098±0.002
YCl_3	0.069±0.002	0.063±0.001	0.05±0.002	0.076±0.012
$YbCl_3$	$0.025 \pm 8 \cdot 10^{-4}$	0.061±0.002	$0.08 \pm 4 \cdot 10^{-4}$	$0.075 \pm 4 \cdot 10^{-4}$

7.1.2 Phase Behavior Under Variation of the Anion

The influence of the anion was investigated, too. Tables 7.3 and 7.4 show the effect on BLG, HSA, BSA, and pepsin induced by yttrium and ytterbium salts with varying anions. Again, a first critical boundary was found in all systems tested.

Table 7.3: *Y-salt systems tested for RC. Systems with c^* and c^{**} are marked by two green checkmarks, systems with only a first boundary by one orange checkmark.* [1]*re-clearing at very high salt concentrations.*

	YCl_3	$Y(NO_3)_3$	$Y_2(SO_4)_3$
BLG	✓✓	✓✓	✓✓[2]
HSA	✓✓	✓	✓
BSA	✓✓	✓	✓
Pepsin	✓[1]	✓[1]	✓

Table 7.4: *Yb-salt systems tested for RC. Systems with c^* and c^{**} are marked by two green checkmarks, systems with only a first boundary by one orange checkmark.* [1]*re-clearing at very high salt concentrations.* [2]*Only partial re-clearing.*

	$YbCl_3$	$Yb(NO_3)_3$	$Yb_2(SO_4)_3$
BLG	✓✓	✓✓	✓✓
HSA	✓✓	✓	✓
BSA	✓✓	✓	✓[2]
Pepsin	✓	✓[1]	✓[1]

The binding of rare earth elements (REE) in complexes to proteins is dominated by electrostatic rather than covalent forces. The complexes are either inner sphere (no intervening solvent molecules) or outer sphere (solvent-separated ion pairs)[217]. While the opinion is divided whether at 25°C in aqueous solutions, SO_4^{2-} and NO_3^- complexes with REE are inner or outer sphere, Cl^- forms predominantly outer-sphere complexes[217]. Chloride and nitrate complexes of rare earth elements are very weak and chloride complexes are the weakest[217]. Our

first assumption therefore was that the strength of the effect of the anions used on the proteins tested follows the inverse order of their complex strength with rare earth elements: $Cl^- > NO^- > SO_4^{2-}$. This classification is confirmed by the observations on BSA and HSA with yttrium and ytterbium salts (Table 7.3 and 7.4): the strong chloride salts, YCl_3 and $YbCl_3$, cause a complete reentrant behavior while the weaker nitrate and sulfate salts only cause a first border in both HSA and BSA. The BSA+$Yb_2(SO_4)$ system only shows a partial re-clearing at higher salt concentrations, reminiscent of *pseudo-c*** in systems of BLG and divalent salts (see publications A-C).

In BLG, a complete RC behavior could be induced with all salts tested. Pepsin samples show no normal reentrant effect. While c^* and c^{**} cannot be induced by a simple pH change in BSA, HSA, and BLG, both condensation and re-clearing are found in pepsin when HCl is added without salts. The transition to turbidity takes place before the isoelectric point is reached but close to it. We therefore assume that the transitions to clear samples at high trivalent salt concentration in pepsin are at least partially caused by the pH change.

7.1.3 Zeta Potential Measurements and Charge Inversion

Zeta potentials of multiple protein–salt systems were measured in order to confirm that charge inversion occurs at high salt concentrations. The zeta potential of systems with a second boundary c^{**} showed typically a dependence on the salt concentration similar to the example of BLG in the presence of $Yb_2(SO_4)_3$ that is shown in Figure 7.1a. The charge inversion is clearly visible. The zeta potentials of systems with no second border or a second border at very high salt concentrations tend to saturate at negative values. However, at much higher salt concentrations, positive zeta potentials are found. Figure 7.1b shows zeta potential measurements of pepsin with YCl_3 as an example. Further plots of other systems can be found in the Appendix E.

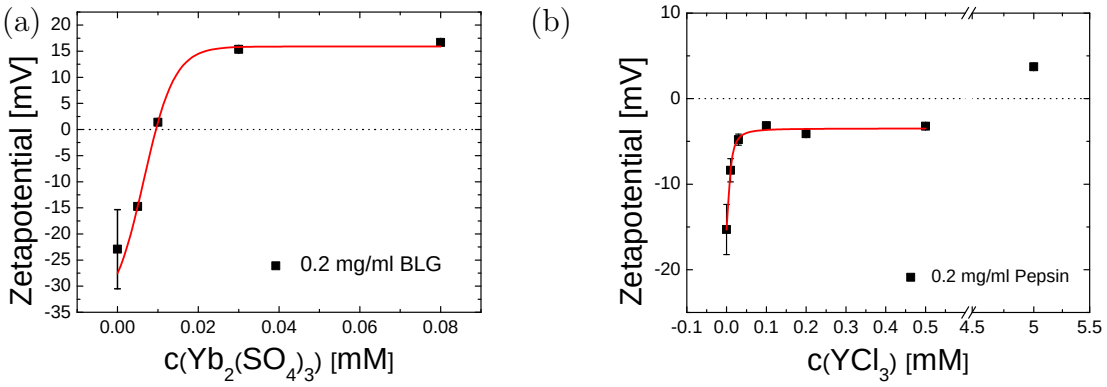

Figure 7.1: *Zeta potential as a function of salt concentration. (a) BLG with $Yb_2(SO_4)_3$; (b) pepsin with YCl_3. The error bars show the standard deviation of five measurements.*

For pepsin, very high c^{**} values were observed with YCl_3 and $Y(NO_3)_3$ (92 and 80 mM at a pepsin concentration of 5 mg/mL - because of its much smaller solubility, no result for $Y_2(SO_4)_3$ could be obtained). They coincide with the eventually positive zeta potentials at high salt concentration. This could be partially due to protonation. However, the pH values of all samples are above the very low isoelectric point of pepsin of 2.2-2.8[178]. The lowest pH value (3.69) was measured for 5 mg/mL pepsin with 80 mM $Y(NO_3)_3$.

7.1.4 Protein Stability

A possible reason for missing second critical borders could be the unfolding and denaturing of the proteins with increasing salt concentrations. Therefore, the stability of the secondary structure of the proteins under the experimental conditions was monitored by FTIR and CD. FTIR and CD measurements of BSA and HSA in the presence of YCl_3 were performed by our group before and no significant changes were found[226]. Likewise, no indications for structural changes are visible in most of the new systems with the following exceptions: The FTIR results show that BSA (20 mg/mL) stays stable from 0 to 10 mM $Yb_2(SO_4)_3$ but decays heavily at 45 mM. Comparing this result to Table 7.4, we see that this system has a first critical border c^* and becomes partially clear without reaching the completely clear state again at high salt concentrations. This behavior is probably caused by unfolding of the protein. CD measurements on this system show no dramatic changes, probably due to the much lower salt concentrations. With $Y(NO_3)_3$ (and in weaker occurrence with $Yb(NO_3)_3$), a structural change in BSA and HSA is probable. Finally, the plots for BLG indicate that the protein has the same structure in the first and third regimes with all salts tested. However, in the second regime, the samples tested show less molar ellipticity than in the first and third. This could be caused by a lower protein concentration due to precipitation and by reversible partial unfolding in the second regime. Plots of FTIR and CD measurements can be found in the Appendix E.

7.1.5 Effective Protein-Protein Interactions Determined by SAXS

In order to further complement our findings about effective interactions in protein solutions tuned by different trivalent salts, SAXS measurements were performed. Figure 7.2 shows measurement curves with 5 mg/mL BSA in the presence of $YbCl_3$ and $Yb(NO_3)_3$. In the case of $YbCl_3$, the effective interactions first change from repulsive to attractive with increasing salt concentration, indicated by the upturn at low q. In the third regime (10 mM $YbCl_3$), the strength of attraction decreases again. In the case of $Yb(NO_3)_3$, the intensity at low q continuously increases. Further measurements are necessary for a better understanding of the salt effects.

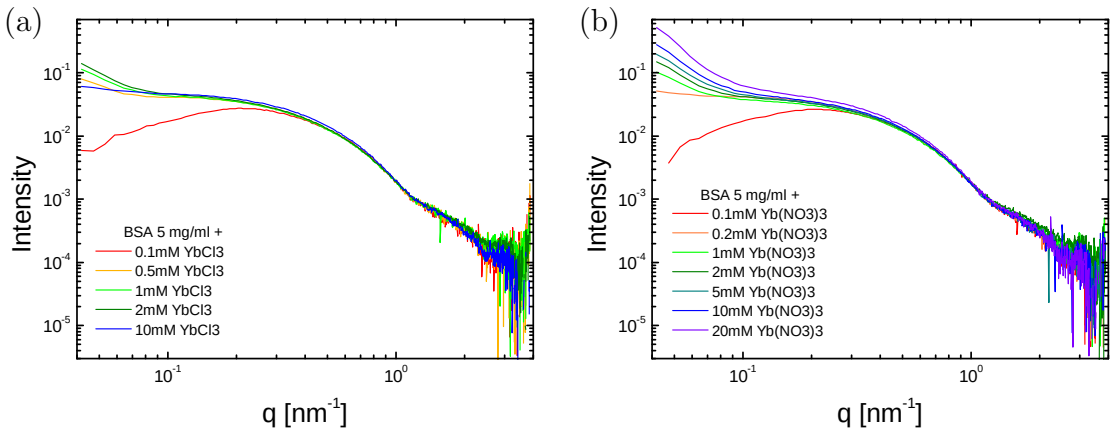

Figure 7.2: *SAXS measurements of 5 mg/mL BSA in the presence of (a) $YbCl_3$, (b) $Yb(NO_3)_3$.*

7.1.6 Crystallization and Structure Determination

One of the central applications of RC in protein systems is tuning the effective interactions between proteins towards crystallization. Applying the experience from the crystallization of BLG with YCl_3, $ZnCl_2$ and $CdCl_2$, samples of the systems presented above were prepared with protein and salt concentrations close to the boundaries, especially c^*, which was found in all systems tested. The samples were stored at 5 and at 21°C. In fact, a large number of the systems tested could be successfully crystallized. Especially crystals of BLG and HSA were found in the presence of various salts: BLG could be crystallized with all salts tested and HSA with all but the three ytterbium salts, including both $Y(NO_3)_3$ and $Y_2(SO_4)_3$, which did not induce a second boundary in HSA. In seldom cases, crystals of BSA and OVA were found; however, the reproducibility is poor and ideal conditions are still to be determined. As an example, Figure 7.3a shows a picture of crystals grown at 21°C from a solution with 40 mg/mL HSA and 4 mM $LaCl_3$ (close to c^{**}) and Figure 7.3b shows a picture of crystals grown at 5°C from

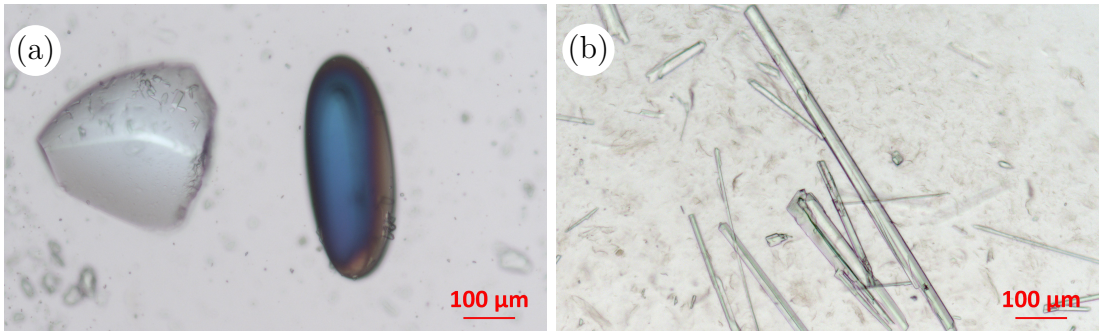

Figure 7.3: *(a) Crystals grown from a solution with 40 mg/mL HSA and 4 mM $LaCl_3$ at 21°C. (b) Crystals grown from a solution with 20 mg/mL BLG and 1.3 mM $Y(NO_3)_3$ at 5°C.*

a solution with 20 mg/mL BLG and 1.3 mM $Y(NO_3)_3$ (close to c^*).

The structure determination by X-ray diffraction is ongoing. Up to now, structural information was collected from BLG in the presence of various salts and of HSA in the presence of YCl_3 and $GdCl_3$. In all of them, crystal contacts are mediated by ion bridging. BLG in the presence of $Y(NO_3)_3$, $Yb_2(SO_4)_3$ and $GdCl_3$ (and possibly also in the presence of $Y_2(SO_4)_3$, where one lattice angle only slightly deviates from 90°) forms, as well as in the already published case of YCl_3 [227], orthorhombic crystals. Table 7.5 summarizes the benchmark data. A detailed discussion of the commonalities and differences in these structures with varying salt type will be subject of future work.

Table 7.5: *Structural information of BLG and HSA crystals determined by X-ray diffraction (work of M. Oelker, S. Da Vela and Dr. G. Zocher). Type and concentration of the protein and the salt, the space group, the cell parameters (lengths a-c and angles α-γ) and the resolution are listed.*

prot.	**$c_{protein}$** [mg/mL]	**salt**	**c_{salt}** [mM]	**space group**	**cell parameters**	**res.** [Å]
BLG	50	$Y_2(SO_4)_3$	1	$P2_1$	a=64.24, b=40.31, c=134.45, α=γ=90°, β=89.72°	2.97
BLG	5	$Y(NO_3)_3$	3	$P2_12_12_1$	a=39.42, b=59.94, c=136.68, α=β=γ=90°	2.11
BLG	10	$Yb_2(SO_4)_3$	0.3	$P2_12_12_1$	a=39.47, b=60.21, c=136.30, α=β=γ=90°	2.35
BLG	20	$GdCl_3$	2	$P2_12_12_1$	a=39.67, b=59.97, c=137.21, α=β=γ=90°	2.70
HSA	31	YCl_3	2	$P2_12_12_1$	a=55.68, b=71.84, c=180.58, α=β=γ=90°	2.11
HSA	30	$GdCl_3$	1.1	$P2_12_12_1$	a=55.06, b=71.31, c=180.1, α=β=γ=90°	3.19

7.2 EpzP - a "Real-World System"

While our research on RC was performed so far on proteins that can be purchased in large amounts, the aim is rather to apply it to the crystallization of proteins of unknown structure which are subject of current biochemical research. The first attempt to induce RC in a not commercially available system involves Phenazine prenyl transferase (EpzP), an enzyme described by Seeger in 2011 and Zocher in 2012 [175,229]. It was chosen because it can be produced in relatively high yields and, moreover, its DNA construct is available which allows performing site-directed mutagenesis. We aim to modify its surface towards more acidic residues in order to allow or optimize its crystallization properties using metal salts. Wild type EpzP (27.3 mg/mL) was titrated with YCl_3 and both c^* and c^{**} were found.

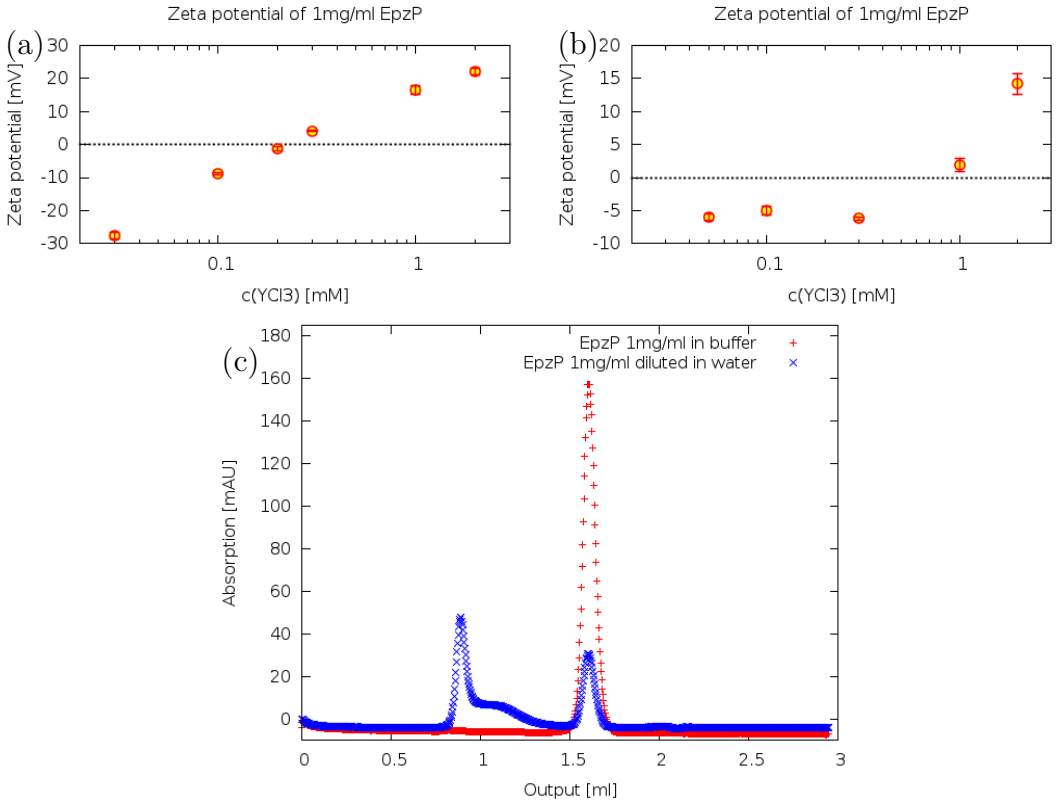

Figure 7.4: *Zeta potential measurements on 5 mg/mL EpzP in dependence on the YCl_3 concentration. (a) with (b) without 5 mM HEPES buffer (pH 7.0). (c) Size exclusion chromatography results obtained by an analytical column (AG Stehle). The error bars show the standard deviation of five measurements.*

Zeta potential measurements on EpzP with YCl_3 in water clearly show a charge inversion, see Figure 7.4a. However, it turned out that EpzP does not remain stable if diluted in pure water. The size exclusion chromatography results presented in Figure 7.4c show a clear shift to aggregates, whereas it stays in its monomeric form if diluted in 5 mM HEPES buffer (pH 7.0). The use of buffer seems to be necessary for this protein. Measurements on the zeta potential on EpzP with YCl_3 in 5 mM HEPES are shown in Figure 7.4b and have an unusual trend, but still indicate charge inversion.

So far, it was not possible to crystallize the wild type using YCl_3. However, the design of the mutant with a tetrahedral acidic patch is ongoing. In the future, it is planned to produce it in larger amounts and to characterize the modified protein. It is planned to crystallize it and compare it to the wild type.

7.3 Discussion

We tested a series of protein–salt systems with regards to cation and anion effects on RC, charge inversion, and RC-induced crystallization. A lower boundary was found in all systems tested.

Variation of the cation has a rather small (and protein dependent) effect, although the effective ion radii are quite different. While BLG, BSA and HSA show a complete RC phase behavior with a second boundary c^{**} in the presence of all the chloride salts tested, pepsin only shows the lower critical boundary in all systems tested. Variation of the anion, however, has a stronger effect: in both HSA and BSA, an upper boundary can only be induced by the chloride salts but not by sulfates and nitrates. In OVA, no second critical boundary was found in the presence of $YbCl_3$ but with other cations. The mechanisms behind this ion-dependence and observations such as the inverse behavior of the slopes of c^* shown in Table 7.2 are not yet understood. Various effects that have not been taken into account up to now, such as Hofmeister effects, differences in the solvent accessible surface of pepsin or BLG and HSA or BSA, the pK_a values of the substances and more should be considered for further interpretation.

One of the central applications of RC in protein systems is tuning the effective interactions between proteins towards crystallization. Indeed, several of the systems tested could be successfully crystallized. Especially crystals of BLG and HSA were found in the presence of various salts.

Earlier expectations were that RC could occur in nearly half of all proteins (that is the negatively charged ones at neutral pH) in the presence of multivalent metal salts[226]. The recent experiments confirm that especially the first boundary c^* is very common among negatively charged globular proteins. A second boundary, however, is not found in all systems tested. Best results were obtained with chloride salts, in accordance with former experiments with $FeCl_3$ and $AlCl_3$[226]. The induction of crystallization was not successful either in all the systems tested but is still a promising approach which could be even further improved with the increasing understanding of the underlying mechanisms. The good crystallization conditions close to the boundaries that were observed in BLG in Chapter A-C and Ref.[227] with various salts were confirmed in additional systems.

Another approach is the targeted modification of proteins towards a better suitability for the crystallization by RC. Proteins with a known structure but poor single crystal diffraction data provide information about the present surface-residues but still have undesirably low resolution. By mutation of the surface-residues, we could design better-diffracting crystals. A first step towards this method is done by the aforementioned experiments with EpzP.

Part III

Conclusions & Outlook

Chapter 8

Summary & Conclusions

In the following, the main findings of this thesis are discussed. Its two main topics are the determination of proper conditions for the nucleation of high quality crystals which is a main challenge for structure biology and to study the early stage of crystallization, including the pathways it follows and its possible precursors. It has been shown that sign and strength of the interactions between negatively charged globular proteins can be modified by salt effects[86,151]. RC could provide a method to promote crystallization in a large number of proteins. In the case of HSA in the presence of YCl_3, crystallization was found in both the dilute and the dense liquid phase after LLPS and in solutions between c^* and the LLPS region[215,216,223]. The main protein used for the work in this thesis was BLG from cow milk.

Earlier experiments with BLG in the presence of YCl_3 suggested that the best conditions for high quality crystals can be found close to the boundaries[227]. This could be confirmed in the systems of BLG in the presence of $ZnCl_2$ (Chapter 3, publication A) and $CdCl_2$ (Chapter 4 and 5, publications B and C). The divalent salts $ZnCl_2$ and $CdCl_2$ cause a normal first boundary c^* in BLG. At high salt concentrations, however, only a partial clearing is found at a concentration we denote as *pseudo-c***, which is in contrast to systems with trivalent salts, where a complete reclearing was observed. A charge-inversion occurs in both cases. Using SAXS, we could show that the effective interactions become less repulsive approaching c^* and finally become attractive. Repulsion is not reached again in the third regime. While crystallization of BLG also takes place deep in the second regime, the quality of the crystals is not sufficient for structure determination. This can be explained by too strong attractive interactions (in accordance with our SAXS measurements) that lead to the formation of stable, noncrystalline aggregates. The metal ions are an integral part of the lattice and the cations form new contacts by ion bridging. While Cd^{2+} and Y^{3+} only bind to acidic side chains, Zn^{2+} also binds to histidine. Crystals grown in the presence of $CdCl_2$ under various tested conditions always have a twinning problem (the structure including the twinning problem has already been reported by A. Green and R. Aschaffenburg in 1959[68]) However, the BLG structures obtained with YCl_3 and $ZnCl_2$ are of high quality and were deposited in the protein data bank.

Which pathways of crystallization are followed is an often discussed question, mainly because energetically favored processes could exist under certain conditions. If these conditions are known, easier ways to obtain crystals can be found.

In Figure 5.1 (page 93), three possible scenarios were presented. In the "classical" pathway, one Gibbs free energy barrier has to be overcome for the formation of a critical nucleus. Nonclassical pathways involve an intermediate phase metastable with respect to the crystalline state. It either can be also metastable with respect to the initial fluid phase (resulting in microscopic clusters) or more stable (enabling the formation of a macroscopic dense liquid phase). However, the presence of an additional phase other than the initial fluid and the crystalline solid does not necessarily imply that these protein clusters or dense liquid phases act as precursors for crystal nucleation. A further distinction has to be drawn into the parallel formation of the additional phase (possibly acting as agent for heterogeneous nucleation and/or as protein reservoir for growing crystals) and the scenario where the crystal nucleation occurs within this precursor. These two cases were shown in Figure 4.1 (page 79), too. We refer to both of them as "nonclassical" and to the latter one as "two-step" or "multistep" (nucleation).

It is likely that close to c^*, crystallization in BLG with all three salts tested occurs via a classical one-step process, although the possibility of nucleation from small clusters cannot be ruled out. Close to (*pseudo-*)c^{**}, additional phases other than the initial solution and the final crystal were observed. This is consistent with findings in literature that specific interactions (patchy interaction) may be important for the formation of a precursor [76,213]. By altering independently the nonspecific and specific interactions, one can influence which crystallization pathway is followed. In the case of our system, this could be selected by variation of the salt concentration: close to c^* the nonspecific interactions might dominate. Increasing the specific interaction mediated by ion-bridges by increase of the salt concentration possibly leads to the non-classical pathways observed in BLG with YCl_3, $ZnCl_2$ or $CdCl_2$. In Chapter 3 (publication A), the systems of BLG in the presence of YCl_3 and $ZnCl_2$ were studied. The solutions are clear (YCl_3) / relatively clear ($ZnCl_2$) and apparently in the third regime above a transition temperature T_{tr} and do not crystallize. Below T_{tr}, two different nonclassical pathways with an additional phase are observed: close to T_{tr}, the additional phase consists of protein clusters or small aggregates. Well below T_{tr}, larger aggregates are found directly after sample preparation which relax into a dense liquid phase within 1-2 h. The initially fine network of dense liquid then coarsens with time. This behavior is known in literature as a special case of LLPS and called viscoelastic phase separation. The phenomenon was described for polymers and for lysozyme [93,191,192]. Crystallization occurs at the interface of this dense and the surrounding dilute liquid phase. However, the observations by optical microscopy are not sufficient to answer the question whether nucleation takes place in two steps in which the dense phase is a direct precursor of the crystal or if it takes place heterogeneously from the dilute phase at the droplet interface ("parallel" pathway).

While it was not possible to decide which one of these nonclassical pathways was followed in the systems discussed in Chapter 3 (publication A), the results shown in Chapters 4-5 (publications B and C) provide evidence for a multistep

process. From observation of the crystallization kinetics, we propose the formation of metastable aggregates with a certain noncrystalline ordering in a first step followed by the nucleation of crystals within this intermediate phase. In the beginning, numerous new crystals appear while their growth is slow which can be explained by a low mobility in the aggregates. After some time (depending on the salt concentration), crystal growth becomes faster, possibly due to contact to the dilute phase after the consumption of enough surrounding aggregates. The dilute phase enables faster diffusion and therefore could explain the faster crystal growth in spite of the lower density. The slower dynamics in the denser phase may be sufficient for crystals to nucleate since less molecules are involved. The observed kinetics can be reproduced well by a rate equation model based on these assumptions. Further experiments on BLG with $CdCl_2$ suggest that increasing the salt concentration in the transition zone *pseudo-c*** leads to fewer or finer distributed aggregates, as well as to fewer (but due the larger amount of available material per crystal larger) crystals.

The system of BLG in the presence of YCl_3 has been further studied under variation of the temperature using a combination of SAXS, SANS and VSANS (very small angle neutron scattering). The considered samples are clear at the preparation temperature of 35°C and are cooled well below their transition temperatures. The scattering data show a hierarchy of structures from monomer to large clusters that changes with decreasing temperature. Real-time SAXS data show isosbestic points, indicating a transformation from one state to another, pursuing equilibrium. The SAXS data can be modeled well by two-state analysis fits. These results could prove to be useful for the understanding of the metastable intermediate phase in this system.

In the end, we want to generalize our work and judge if our findings can be transferred to other systems. A systematic study with various proteins and trivalent salts was performed. In all of these systems, a first boundary could be induced and in most cases also a second one. In HSA and BSA, an upper boundary only can be induced by the chloride salts but not by sulfates and nitrates. The cation also has a certain effect: in OVA, no second boundary was found in the presence of $YbCl_3$, but with other chloride salts. The reasons behind these effects, however, are not completely understood yet. BLG and HSA could be crystallized in the presence of many different trivalent metal salts. The structure determination by X-ray diffraction has successfully been performed for some of the systems and is ongoing in others. The induction of RC by multivalent salts thus is a promising approach for the crystallization of protein systems. Protein and salt concentrations close to the boundaries are especially suitable for the growth of high quality single crystals. In all determined structures, crystal contacts are mediated by ion bridging. The aimed modification of proteins towards a better suitability for our crystallization method offers the possibility to design better-diffracting crystals. A first step towards this aim is done by our experiments with EpzP.

Concluding, it can be stated that the tuning of protein–protein interactions by the addition of multivalent metal salts provides a promising method for the growth

of high-quality crystals. Our real-time studies provide a noninvasive method to follow the kinetics of crystallization and comprehend the underlying nucleation mechanisms.

Chapter 9

Remarks, Problems, and Outlook on Future Work

In this thesis, several studies of aggregation and especially crystallization of globular proteins have been presented and the possible pathways of nucleation and growth have been discussed. Collating the results of the individual chapters, a consistent picture could be established. Naturally, there is still some improvement possible. In this chapter, open questions and experimental difficulties will be addressed and future perspectives discussed.

The analysis of the visible area of crystals by optical microscopy is limited by the resolution and — especially for increasing automation — raises the problem of blurred crystals from other layers being taken into account. The SAXS analysis is hindered by crystals or dense liquid phase falling out of or inside the beam due to gravity. This hindered closer investigations of the crystallization pathway of BLG in the presence of $ZnCl_2$ via a liquid network: the denser material sinks to the bottom of the SAXS capillary before crystallization starts. In the system with $CdCl_2$, especially within the *pseudo-c*** zone, samples are fortunately rather viscous and stabilize the crystals at their positions. Only at the end of the growth process, probably partially due to the consumption of the viscous material and partially due to the increasing weight of the crystals, the crystals sediment. For the future, a possibility could be the preparation in a hydrogel, as performed e.g. by Finet *et al.*[48]. However, this probably also will change the phase behavior. As shown in Chapter 5 (publication C), the Bragg peaks are very pronounced in SANS measurements in D_2O, too. All features relevant for our analysis can be measured at the smaller sample-to-detector position (1.5 m) where the experiment can be performed fast enough to enable real-time measurements of the crystallization process with a temporal resolution similar to the one applied with SAXS in Chapters 4-5 (publications B and C). Real-time SANS could eliminate two problems of the SAXS measurements: possible beam-damage by too long exposure time and, due to the larger beam size of SANS, difficulties in collecting data for samples at higher salt concentrations, where the number of crystals decreases significantly (see also Chapter 5 / publication C).

The effect of the vial surface should not be disregarded. In the system of BLG and $ZnCl_2$, similar networks of dense liquid form at the bottom of hanging drops (sample–air interface) as well as on the walls of glass capillaries. Therefore it can be concluded that they are not due to wetting on the glass–liquid interface.

However, in other systems, an influence of the vial material could be observed: for the crystallization of BLG in the presence of $CdCl_2$ followed by optical microscopy, usually hydrophobically coated glass slides were used. Exchanging them by normal glass, noticeably more crystals are formed and the different growth speeds (the "kink" in the average length in Figure 4.4g, page 83) is not observable anymore. This could be connected to enhanced heterogeneous nucleation at the glass interface. The impact is even more drastic in e.g. the system of HSA in the presence of YCl_3, where crystallization could be observed in glass and plastic tubes and quartz cuvettes only, but not on hydrophobically coated glass slides or likewise hydrophobically coated (plastic) sitting drop plates. Furthermore, the influence of the solvent (H_2O or D_2O) changes the phase behavior differently strong in different systems. While Braun *et al.* observe that the occurrence of LLPS depends on whether H_2O or D_2O are used as a solvent [25], in the work for this thesis, no crystallization of BLG in the presence of YCl_3 was found in D_2O. With $CdCl_2$ and $ZnCl_2$, however, the crystallization ability of BLG was as good as in H_2O. The reasons behind these solvent dependencies are not fully understood yet and might be connected to the weakened strength of hydrogen bonds in heavy water.

While with $ZnCl_2$ and YCl_3, BLG was observed to form a dense liquid network at certain temperatures and concentrations, no LLPS was found in the presence of $CdCl_2$. The reason of this salt-dependent difference is unknown, too. Naturally, a LLPS region outside the investigated phase diagram region is possible.

The real-time studies presented in this thesis can be extended to other systems in the future. Promising systems are HSA crystallizing in the presence of $LaCl_3$ or $CeCl_3$. These systems crystallize within hours in contrast to several other tested systems that crystallize much slower, in the order of days. Slower crystallizing samples, however, could be potentially measured at a X-ray home-source. On the strength of experience, it is more probable to observe a two-step nucleation process close to the upper RC boundary. Future research can potentially elucidate the question whether the multistep pathway of crystal nucleation that was observed for BLG in the presence of $CdCl_2$ is a frequent property in protein systems or rather an exception. Based on the findings in this field during the last years, it seems to be possible that different two-step or multistep pathways exist depending on the systems and sample conditions. While in Section 1.2, a two-step pathway proposed by ten Wolde, Vekilov and others is presented that suggests a density change followed by a structure change, we observe a rather structural development: the aggregates in our solutions of BLG and $CdCl_2$ form very early and are visible already in the first microscope pictures. We did not follow their formation up to now. In our SAXS study, we observe the subsequent formation of a — not yet crystalline — precursor structure that we assume to be connected to typical length scales inside the aggregates. This precursor subsequently transforms into the crystalline state. In order to gain a deeper understanding of this mechanism, it would be useful to find suitable methods to study the morphology of the MIP more closely. Within *pseudo-c***, the SAXS curves follow a power law with an

exponent slightly below 2 at low q and deviate strongly from the form factor of BLG. Therefore one can conclude that structures larger than the accessible q-range are dominating, which is consistent with the observation of large protein clusters under the optical microscope. The power law indicates a fractal structure. Further information is needed in order to judge how these aggregates further develop and where the typical length that shows in the SAXS curves as a broad peak is located.

Pursuing the first experiments with EpzP, proteins could be modified in order to grow better-diffracting crystals in the future.

Appendix

Appendix A

Supporting Information for Publication A

Figure A.1 shows a hanging drop experiment with top droplet-glass and the bottom droplet-air interface in focus. Only insular dense liquid spots and later crystals are visible on the upper glass surface while a liquid network forms at the bottom of the droplets.

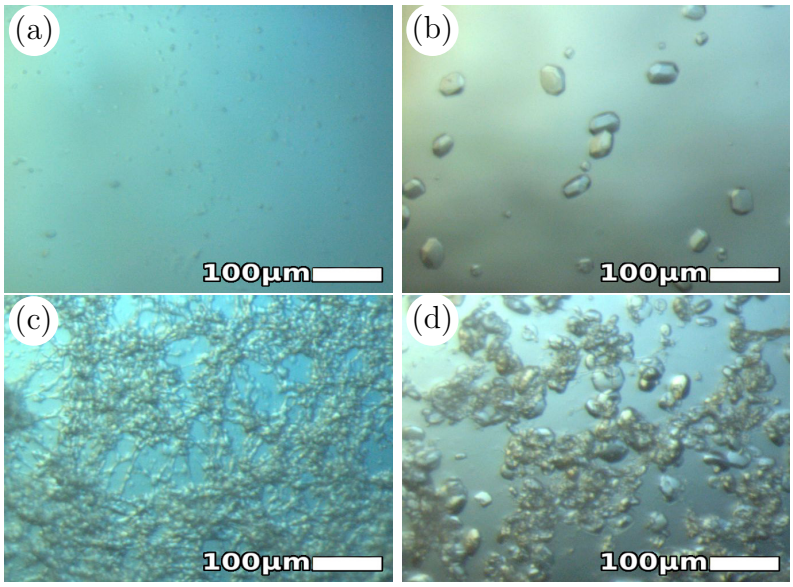

Figure A.1: *Hanging drop experiment of a sample with 6.5 mg/mL BLG and 10 mM $ZnCl_2$ with top droplet-glass (a&b) and the bottom droplet-air interface of the same droplet (c&d) in focus. Only insular dense liquid spots and later crystals are visible on the upper glass surface while a liquid network forms at the bottom of the droplets. a&c: 2 h after preparation, b&d: 23.5 h after preparation*

In Figure A.2, the crystallization process via dense liquid network followed by optical microscope is shown for a sample with 65 mg/mL BLG and 13 mM YCl_3 at 293 K.

Table A.1 shows the fitting parameters obtained for 3.3 mg/mL BLG with varying $ZnCl_2$ concentrations, fitted by an elliptic form factor. The scattering length density (SLD) of the ellipsoid was fixed to the calculated value.

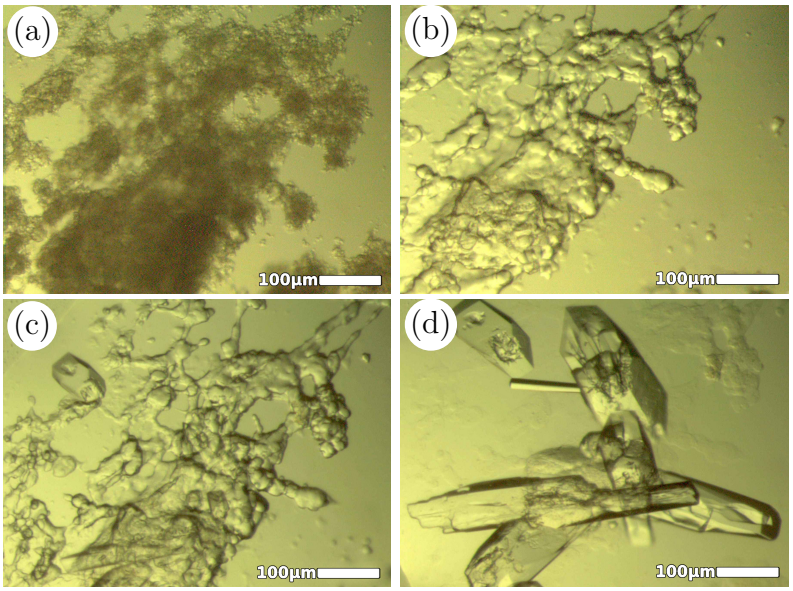

Figure A.2: *Crystal growth at 293 K and a protein concentration of 65 mg/mL with 13 mM YCl_3. The hanging drop method was applied without prior centrifuging. (a) 1 h, (b) 17 h, (c) 45 h and (d) 144 h after preparation.*

Table A.1: *Fitting parameters of the SAXS analysis with a BLG concentration of 3.3 mg/mL.*

c($ZnCl_2$) [mM]	scale [10^{-4}]	a (rotation axis) [Å]	b [Å]
0.1	2.4	37.6	19.2
0.3	2.5	42.6	19.7
20.0	2.8	18.0	55.8
50.0	2.8	18.5	57.4

Figure A.3 shows SAXS data for 65 mg/mL BLG in the first (a) and in the third regime (b) as well as the respective fit curves.

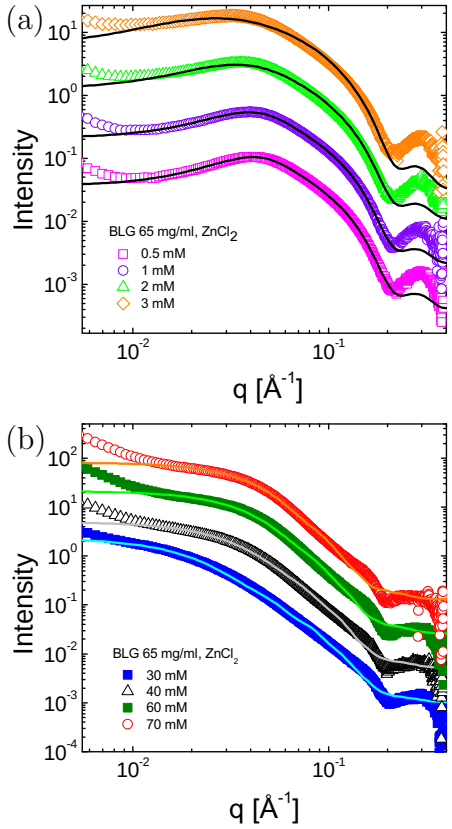

Figure A.3: *Fitted combination of form and structure factors for 65 mg/mL BLG in the first (a) and in the third regime (b). Deviations of the fit from the data at high q values are caused by deviations of the ellipsoid used as form factor from the true form of the BLG dimer.*

SAXS data in Regime I were fitted using an elliptic form factor with screened Coulomb structure factor which was developed by Hayter and Penfold[73,77]:

$$U_{SC}(r) = \begin{cases} \frac{Z^2 e^2}{\epsilon(1+\kappa_D R)^2} \frac{exp(-\kappa_D(r-2R))}{r} & r > 2R \\ \infty & r \leq 2R \end{cases} \tag{A.1}$$

with Z the protein surface charge, e the electronic charge, and ϵ the dielectric constant of the solvent. κ_D is the inverse of the Debye screening length, determined by the ionic strength of the solution and R the effective sphere radius of the ellipsoid with the axes a, b, b. First the data obtained with 0.5 mM $ZnCl_2$ were fitted and the values of the axes a, b fixed to the obtained values for the fits for higher salt concentrations.

In the third regime, an elliptic form factor with a sticky hard sphere structure

factor was used. The sticky hard sphere potential is defined by[15]

$$U_{SHS}(r) = \begin{cases} \infty & r < \sigma \\ -u & \sigma < r < \sigma + \Delta \\ 0 & r > \sigma + \Delta. \end{cases} \tag{A.2}$$

The stickiness parameter τ is defined as

$$\tau = \frac{1}{12\epsilon_\Delta} exp(-u/kT) \tag{A.3}$$

with the preturbation parameter ϵ_Δ

$$\epsilon_\Delta = \Delta/(\sigma + \Delta) \tag{A.4}$$

(fixed to 0.01). As visible by the increase of τ with increasing salt concentration, the interaction strength becomes weaker.

σ is the hard sphere diameter (σ=2R), Δ is the width of the square well, and u is the depth[90]. The obtained volume fractions are close to the calculated value of 0.0488, using the specific volume of BLG of 0.751 l/kg[118]. The dielectric constant was fixed to 81 and the temperature to 298. In all fits, the scattering length density (SLD) of the ellipsoid was fixed to the calculated SLD of BLG for X-rays with the wavelength used. For the SLD of the solvent, always a value between $1.14\,10^{-5}$ Å^{-2} and $1.15\,10^{-5}$ Å^{-2} (124600 eV) or between $1.12\,10^{-5}$ Å^{-2} and $1.13\,10^{-5}$ Å^{-2} (16000 eV) was obtained.

All fitting parameters obtained from Figure A.3 for 65 mg/mL BLG are listed in Table A.2

Table A.2: *Fitting parameters of the SAXS analysis with a BLG concentration of 65 mg/mL.*

c($ZnCl_2$) [mM]	vol. fract.	a [Å]	b [Å]	charge	τ
0.5	0.050	75.5	18.7	9.8	-
1	0.045	75.5	18.7	9.5	-
2	0.045	75.5	18.7	6.1	-
3	0.045	75.5	18.7	3.5	-
30	0.040	19.0	116.0	-	0.19
40	0.055	20.1	80.8	-	0.27
60	0.055	21.2	72.9	-	0.35
70	0.040	21.8	66.6	-	0.33

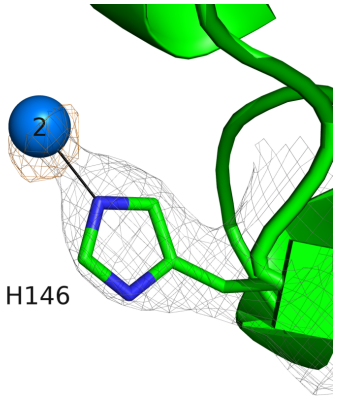

Figure A.4: *Additional Zn binding site at H146 in BLG-Zn-2. This binding site is occupied concentration dependently and was refined to possess an occupancy of 30% for BLG-Zn-2 and 75% for BLG-Zn-20, respectively. Density maps were calculated with FFT[195]. The ($2F_{obs}$-F_{calc})-electron density map (gray) and anomalous difference map (orange) are contoured at σ = 1.0 and σ = 4.0, respectively.*

Table A.3: *Data collection statistics. Values in parentheses are for the highest resolution shell.*

	BLG + 2 mM $ZnCl_2$	**BLG + 20 mM $ZnCl_2$**
Beamline	SLS X06DA	PETRA III P13
Detector	PILATUS 2M-F	PILATUS 6M-F
Wavelength [Å]	1.0000	1.2380
Space group	$P3_221$	$P3_221$
Unit cell [Å]	a = b = 92.83, c = 42.31	a = b = 89.82, c = 40.75
Matthews coefficient [$Å^3$/Da]	2.88	2.59
Solvent content [%]	57.23	44.21
Molecules/ASU	1	1
Resolution range [Å]	19.73 - 2.40 (2.46 - 2.40)	38.89 - 2.44 (2.50 - 2.44)
No. reflections		
Measured	180247 (11948)	114034 (7284)
Unique	15959 (1211)	13636 (939)
R_{meas} [%]	4.5 (103.0)	4.8 (14.3)
Completeness [%]	99.8 (99.5)	99.4 (93.9)
Redundancy	11.3 (9.9)	8.4 (7.8)
$<I>/<\sigma(I)>$	29.62 (3.05)	30.57 (14.86)
Wilson B-factor [$Å^2$]	71.0	59.8
Mosaicity [°]	0.16	0.20

Table A.4: *Refinement statistics.*

	BLG + 2 mM $ZnCl_2$	**BLG + 20 mM $ZnCl_2$**
Resolution range [Å]	19.73 - 2.40	38.89 - 2.44
R_{cryst}	0.1960	0.2037
R_{free} (test set of 12 %)	0.2315	0.2332
No. of non-H atoms		
Protein	1261	1263
Metal ions	2	3
Water	4	10
Average isotropic B-factor [Å^2]		
Main chain	71.9	53.4
Side chain	78.1	57.9
Metal ions	70.3	63.6
Water	63.0	54.1
rmsd bond lengths [Å]	0.004	0.004
rmsd bond angles [°]	0.898	0.871
Ramachandran regions		
Most favorable [%]	96.8	95.5
Allowed [%]	1.3	3.3
Outliers [%]	1.9	1.3

Appendix B

Supporting Information for Publication B

Video on crystallization

The crystal growth and the number of crystals observed by optical microscopy in a time series can also be followed in the video movie.mpg attached as separate file.

FTIR and Zeta-potential data on protein stability

Fourier transform infrared spectroscopy (FTIR) measurements in D_2O and circular dichroism (CD) measurements in H_2O were performed to prove that BLG stays in its native state when increasing concentrations of $CdCl_2$ were added (main manuscript). Here, in Figure B.1a, we show FTIR measurements over a time span of 85 min for the sample condition mainly used in this work (20 mg/mL BLG, 15 mM $CdCl_2$) to demonstrate that there are also no time-dependent changes in protein stability. Circular dichroism requires low protein concentrations at which no turbidity in regime II is visible. To confirm that the complete charge inversion takes place in the investigated $CdCl_2$ range, zeta potential measurements were performed (Figure B.1b).

Additional SAXS measurements and analysis

In Figure B.2, the position of the minimum intensity and the exponent of the intensity as a function of q, $I(q) \propto q^{\alpha}$, in the low q region are plotted as a function of time. The q values at the minimum intensity decrease slightly before crystallization, followed by an increase after 30 min, which is also the moment at which Bragg peaks become detectable as illustrated by the dashed black line in Figure B.2. The slope of the scattering intensity in low q changes from $q^{-2.0}$ to $q^{-2.7}$, indicating the formation of more compact structures.

Figures B.3-B.5 show additional real-time SAXS data and analysis for a different sample within the region of *pseudo-c*** (33 mg/mL BLG with 17 mM $CdCl_2$).

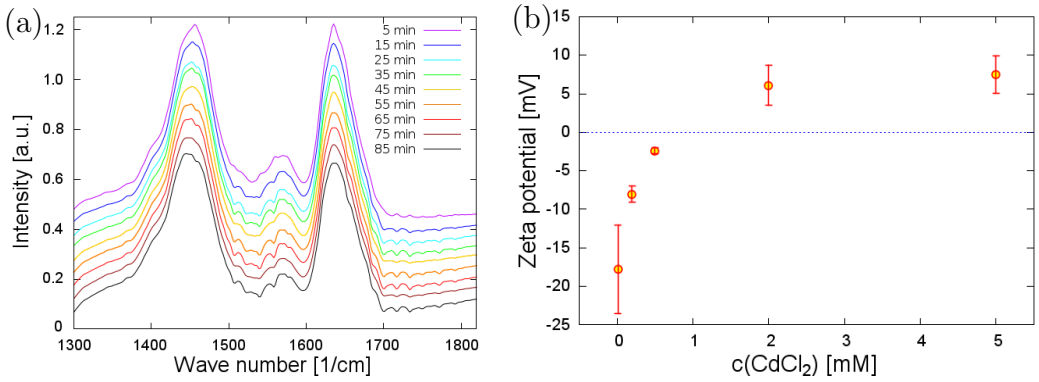

Figure B.1: *(a) Time-resolved FTIR spectra for a BLG concentration of 20 mg/mL and a $CdCl_2$ concentrations of 15 mM covering the time span of the SAXS experiments in time steps of 5 min. Curves are shifted in intensity for better visibility. (b) Zeta potential measurements for samples used for the CD experiments.*

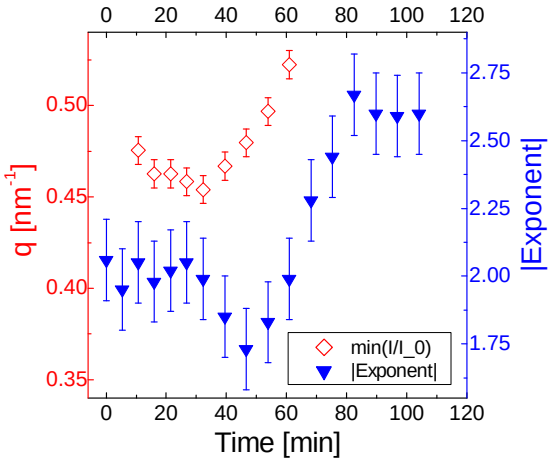

Figure B.2: *Position of scattering minimum of intermediate structure in $I(q)/I_0(q)$ and absolute value of exponent at low q for 20 mg/mL BLG with 15 mM $CdCl_2$.*

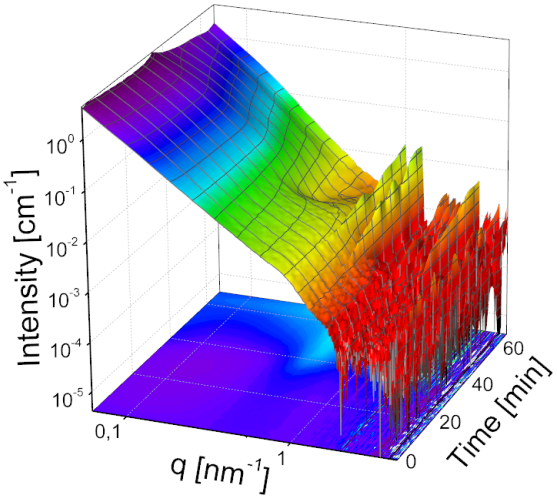

Figure B.3: *3D-plot 33 mg/mL BLG + 17 mM $CdCl_2$.*

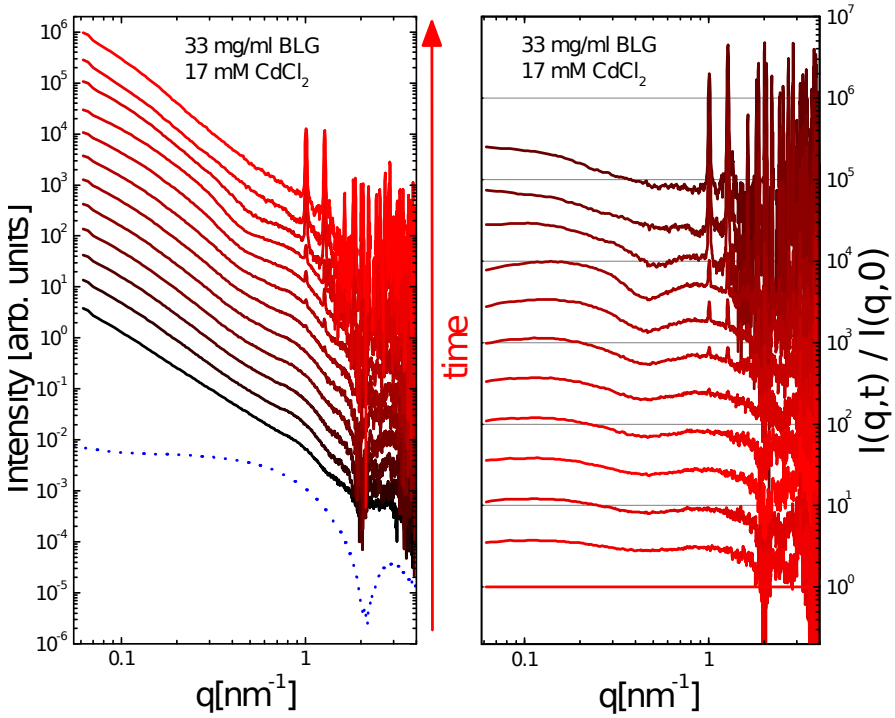

Figure B.4: *Overview: 33 mg/mL BLG, 17 mM $CdCl_2$.*

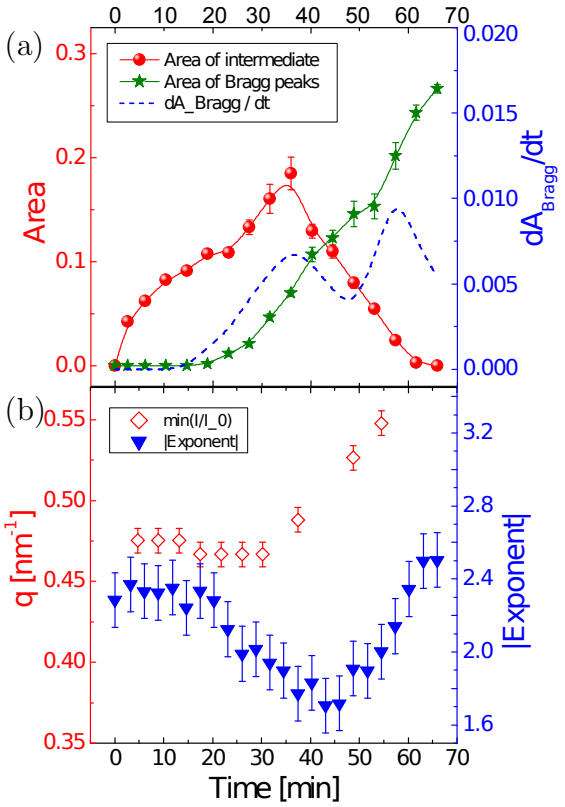

Figure B.5: *Time-dependent development of $I(q)/I_0(q)$ in 33 mg/mL BLG with 17 mM $CdCl_2$. (a) Red spheres (left axis) show $A_{interm}(t)$, the integral of the broad Gaussian function connected to the intermediate. Green stars (left axis) show $A_{Bragg}(t)$, the integral of the two Bragg peaks in this area. For clarity and better statistics, always three data points of $A_{interm}(t)$ and $A_{Bragg}(t)$ were merged into one. The blue dashed line (right axis) shows the time derivative of $A_{Bragg}(t)$, the crystallization rate. (b) Position of minimum of intermediate structure and absolute value of exponent at low q. For clarity, only every third (minimum position) or second (exponent) data point is shown.*

Appendix C

Supporting Information for Publication C

The electrophoretic mobility was used to determine the zeta potential of proteins as a function of the salt concentration. By this method, charge inversion was verified. The measurements were done using a Zetasizer Nano ZS from Malvern Instruments Ltd with a 633 nm laser. Although the reentrant effect for BLG solutions in the presence of $CdCl_2$ is not complete, a charge inversion with increasing divalent salt concentration is observed, see Fig. C.1. Compared to previous work[227], the charge inversion of BLG is weaker and occurs at higher salt concentrations than with the trivalent salt YCl_3[158,227]. The phase boundaries are shifted towards higher salt concentrations, too. This relative weak charge inversion may explain the phase behavior observed, i.e. a $pseudo - c^{**}$ or a transition zone and the sample solutions in the third regime are still a bit turbid.

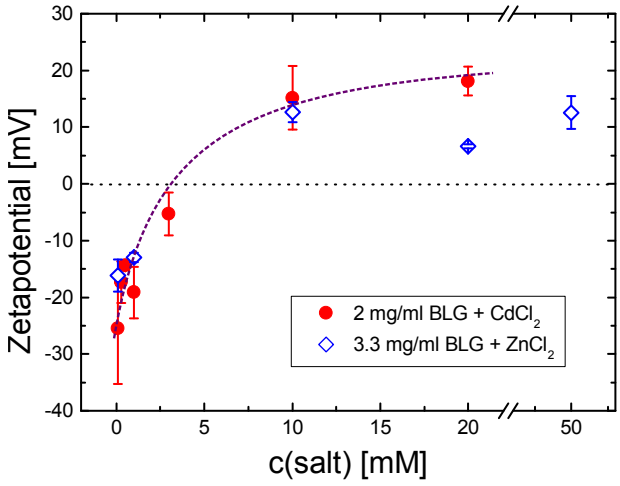

Figure C.1: *Zeta potential measurements on 2 mg/mL BLG with increasing $CdCl_2$ concentrations. The dotted red line is a guide to the eye, the dashed black line marks a charge of 0.*

Protein crystallization followed by real-time SAXS measurements. Samples contain 33 mg/mL BLG with 15 - 20 mM $CdCl_2$ in the transition zone of *pseudo-c***. Fig. C.2 shows additional SAXS data for 33 mg/mL BLG with 15 and 18.5 mM $CdCl_2$.

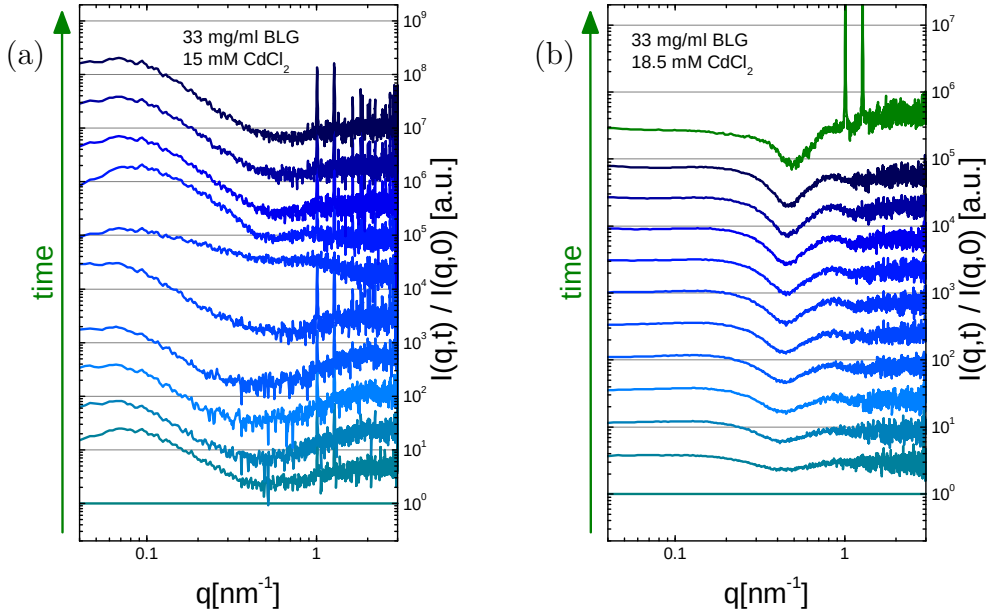

Figure C.2: *Real-time SAXS curves for 15 mM $CdCl_2$ (a) and 18.5 mM (b).*

Fig. C.3 shows the kinetic evolution for a two-step process as modeled by rate equations with parameters reproducing the experimental findings from Figure 8b in the main manuscript.

The number of crystals in solutions with 33 mg/mL BLG and different $CdCl_2$ concentrations around $pseudo - c^{**}$ observed by microscopy can also be followed in the videos listed in the following:

- The first video shows the crystallization in presence of 17 mM $CdCl_2$ (file is attached in electronic supporting information). The crystals are numerous and small.
- The second one shows the crystallization in presence of 20 mM $CdCl_2$ (file is attached in electronic supporting information). Compared to the previous video, significantly fewer and larger crystals can be observed.

Another animation provided with this publication shows the fitting performed for evaluation of the amount of MIP and crystalline phase for a sample with 17 mM $CdCl_2$ (file is attached in electronic supporting information). Both the broad peak and the Bragg peaks have been fitted by scaled Gaussians and their enclosed areas evaluated.

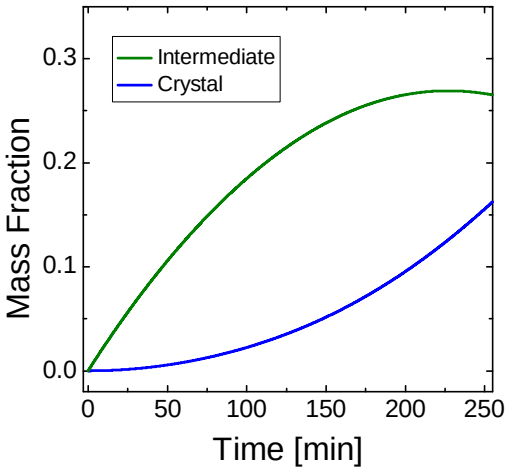

Figure C.3: *Kinetic evolution for a two-step process as modeled by rate equations. The following parameter values were used to show the good qualitative agreement of the model with the data set:* $k_e = 0.15\,min^{-1}$, $L_0 = 0.2$, $\alpha_I = 0.2$, $k_{gL} = 0.6\,min^{-1}$ *(same as for Figure 9c in the main manuscript),* $k_I = 0.003\,min^{-1}$, $k_n = 0.002\,min^{-1}$, $k_{gI} = 0.02\,min^{-1}$.

Appendix D

Supporting Information for Publication D

Table D.1: *Parameters for conventional fitting of involved structures at different length scales in samples with 67 mg/ml BLG and YCl_3 concentrations from 12 to 15 mM (first measurement curve at 35°C and last measurement curve at 5°C). The SLD of the sample was fixed to $1.68 \cdot 10^{-6} Å^{-2}$, the calculated value for BLG for neutrons of the used wavelength. The applied form factors were spheres. For the dimer, a hard sphere structure factor was used with a spherical form factor (radius fixed to 15 Å). In case of the cluster, elliptical form factors were assumed. The curve for 12 mM YCl_3 was fitted by a sticky hard sphere structure factor (perturbation parameter fixed to 0.01), the curve for 13 mM with a hard sphere structure factor and the curves for 14 and 15 mM YCl_3 with a screened Coulombic potential. Ionic strength, temperature and dielectric constant were fixed.*

	12 mM	**13 mM**	**14 mM**	**15 mM**
Dimer 35°C	HS	HS	HS	HS
volume fraction	0.229	0.225	0.224	0.222
Cluster 35°C	SHS	HS	SC	SC
rotation axis [Å]	35.8	32.9	32.6	30.1
non-rotation axis [Å]	93.4	86.4	89.4	90.9
volume fraction	0.054	0.038	0.069	0.095
stickiness	0.2	-	-	-
charge	-	-	0.12	4.95
Cluster 5°C	SHS	HS	SC	SC
rotation axis [Å]	50.1	44.1	44.8	42.6
non-rotation axis [Å]	97.4	98.5	105.6	106.9
volume fraction	0.025	0.101	0.132	0.175
stickiness	0.2	-	-	-
charge	-	-	0.18	4.94

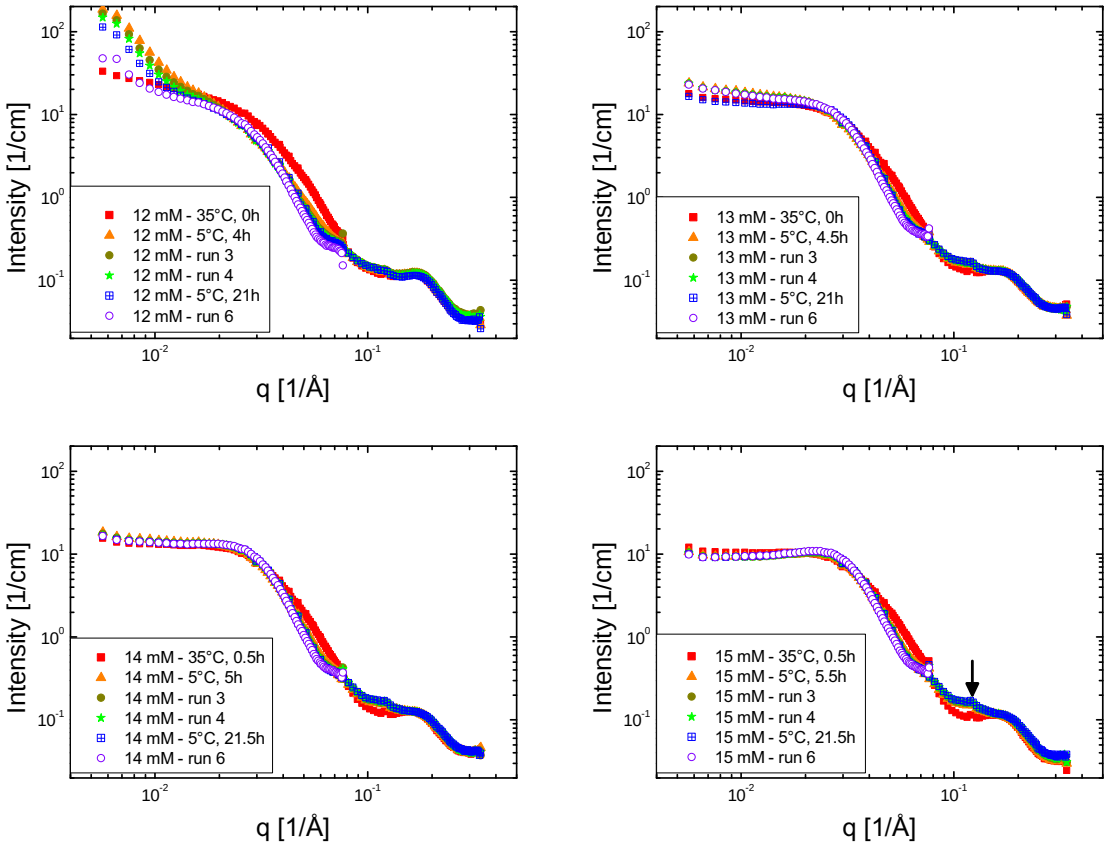

Figure D.1: *Complete set of measurements for Figure 2.*

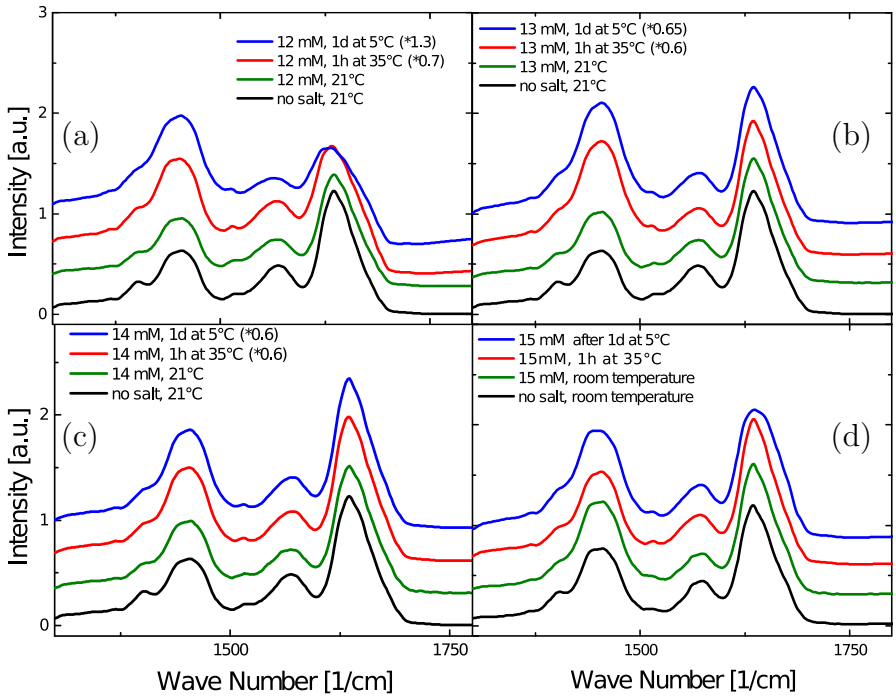

Figure D.2: *FTIR measurements of protein samples with a BLG concentration of 67 mg/mL without salt at room temperature (black) and with (a) 12 mM, (b) 13 mM, (c) 14 mM, and (d) 15 mM YCl_3 at room temperature (green), after preparation and 1 h equilibration time at 35-38 °C (red), and red marked sample after one day at 5 °C (blue).*

Table D.2: *Fitting parameters obtained by a two level Beaucage fit for the first (noncryst-1) and the last measurement run of the deeply quenched sample (noncryst-2) as well as the first (cryst-2) and the final curve of the weakly quenched samples from Ref. [221] (cryst-2). The scaling factor was 0.36-0.38 for the measurements of the deep quench and 0.39 (start) and 0.40 (end) for the crystallizing sample, the background was taken into account with 1.2-2.4·$10^{-3}cm^{-1}$.*

	noncryst-1	**noncryst-2**	**cryst-1**	**cryst-2**
Level 1				
R_g [Å]	130.22	167.03	134.10	138.09
Power	2.67	2.44	2.57	2.51
Level 2				
R_g [Å]	25.5	23.7	24.8	26.8
Power	3.31	3.03	3.79	3.27

Appendix E

Appendix to Chapter 7

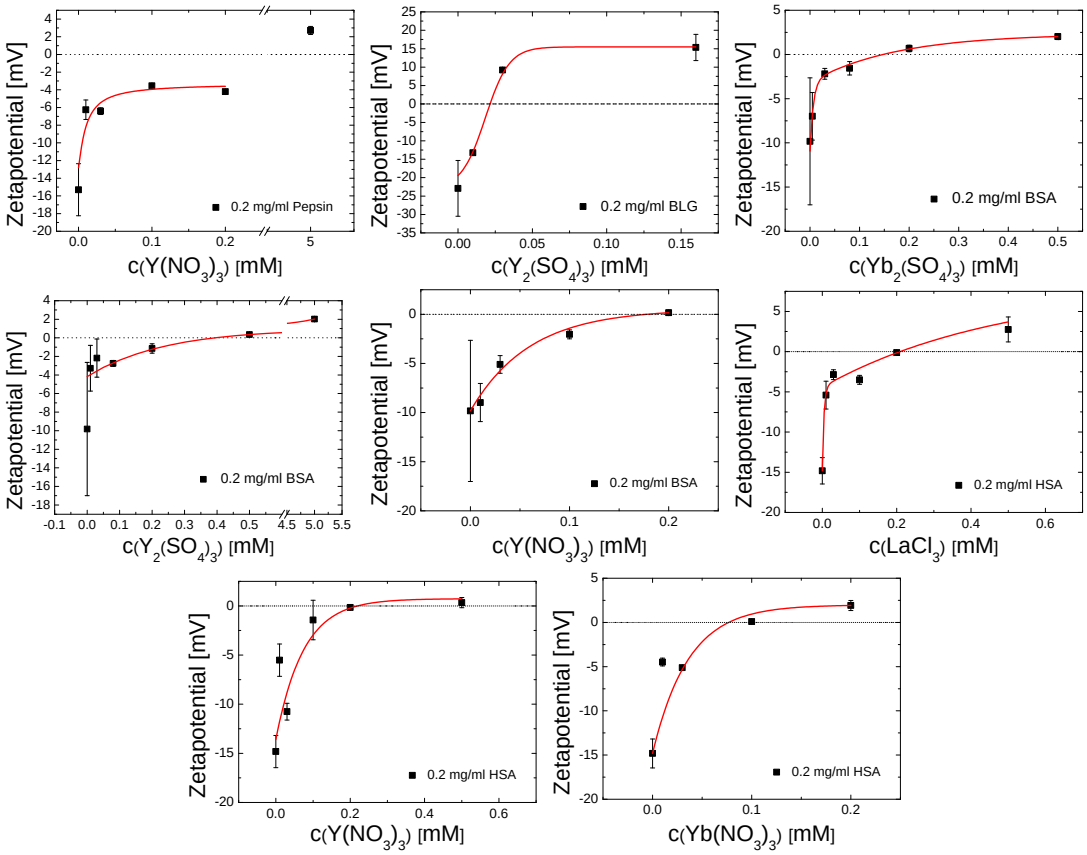

Figure E.1: *Zeta potential measurements of different protein–salt systems.*

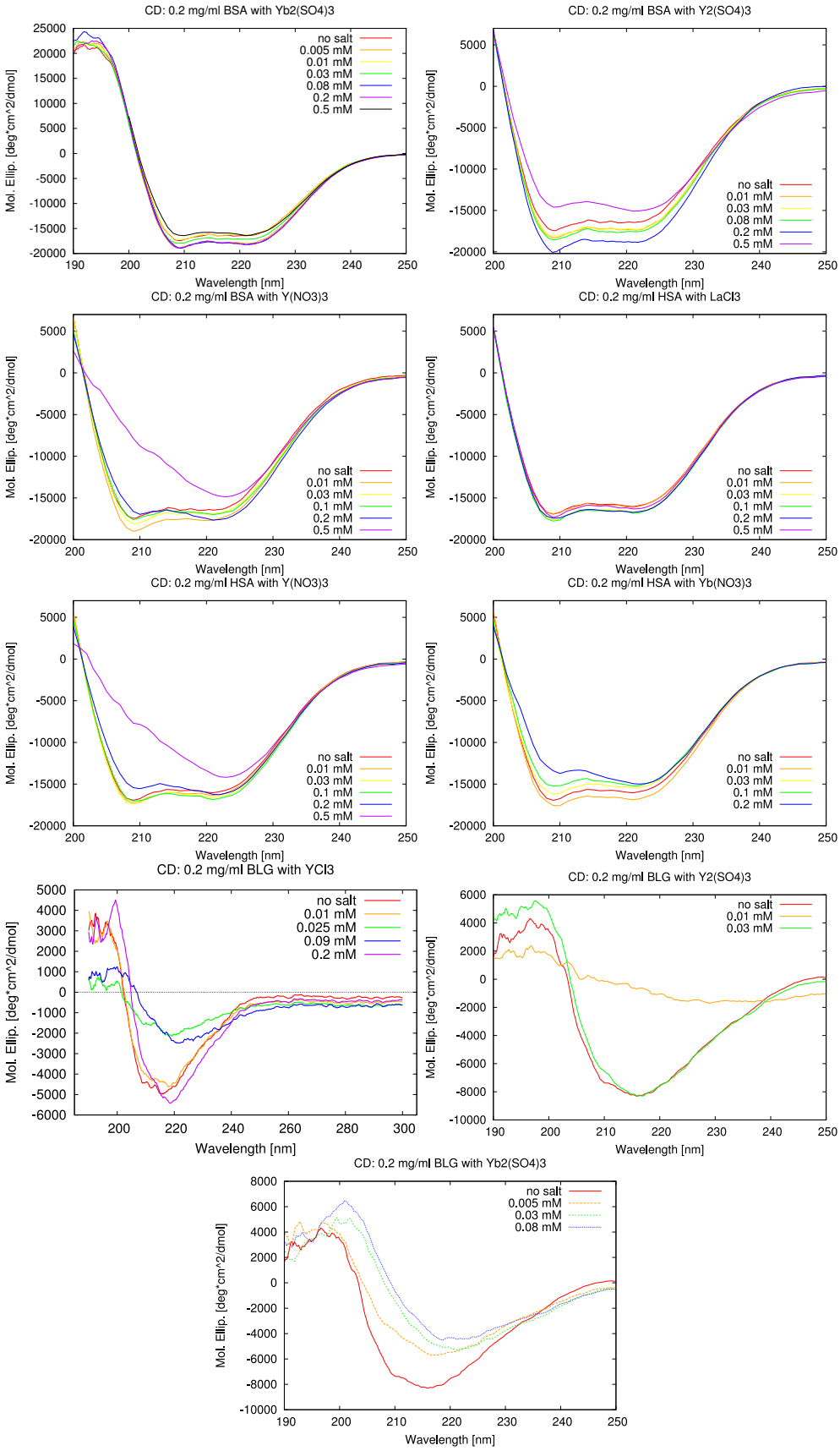

Figure E.2: *Circular dichroism measurements of BSA (top), HSA (middle) and BLG (bottom) with different salts.*

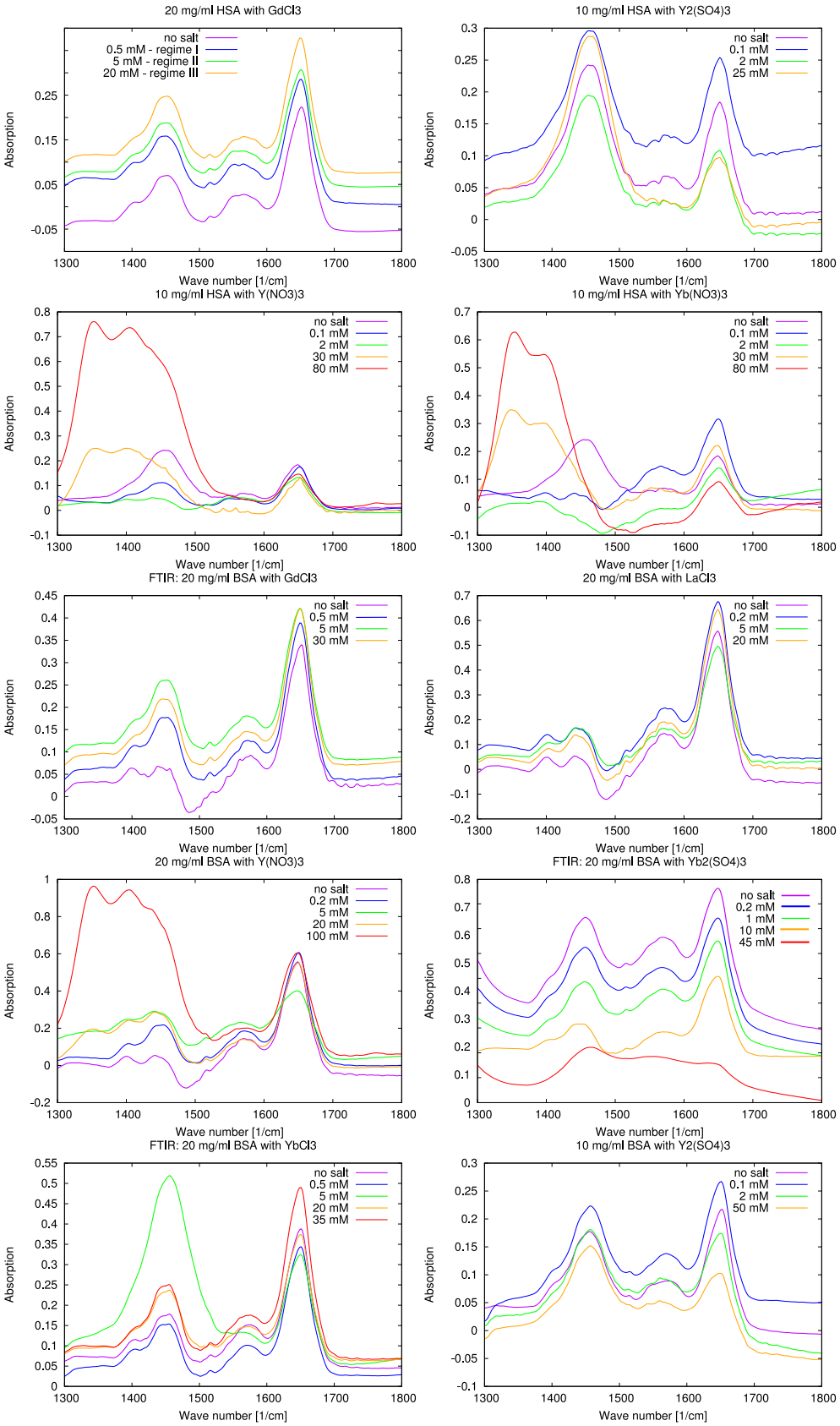

Figure E.3: *FTIR measurements of different protein–salt systems. (I)*

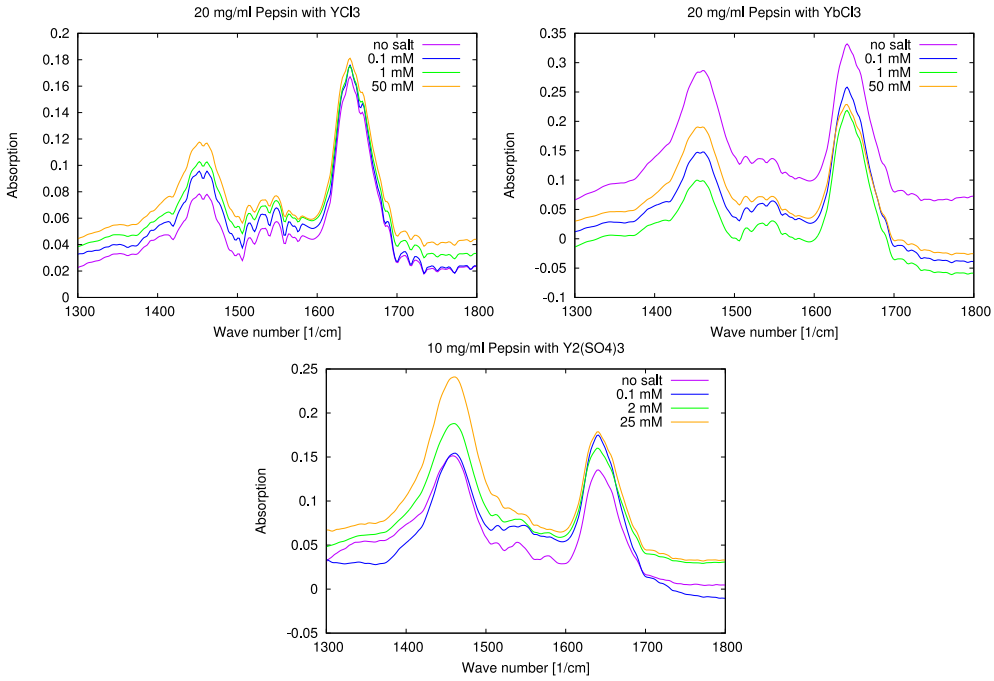

Figure E.3: *FTIR measurements of different protein–salt systems. (II)*

Symbols and Abbreviations

2θ	scattering angle
A	area
$\mathcal{A}$	scattering amplitude
A_2	second virial coefficient
a	pressure resulting from interactions between particles of a gas
$\tilde{a}$	molar a
b	excluded volume contribution in a van der Waals gas
$\tilde{b}$	molar b
b_j	scattering length
c	concentration
c^*	first boundary concentration in RC systems
c^{**}	second boundary concentration in RC systems
D	size in real space
d	sample thickness
$\mathscr{d}$	fractal dimension (Porod exponent)
$d\sigma/d\Omega$	differential scattering cross section
ΔG	change in Gibbs free energy
$\Delta\mu$	change in chemical potential
$\Delta\rho$	difference in scattering length density
$\Delta\Omega$	solid angle
ϵ	extinction coefficient
F	(Helmholtz) free energy
G	free enthalpy or Gibbs free energy
$g(\vec{r})$	pair correlation function
γ	interfacial free energy density
H	enthalpy
$h;k;l$	Miller indices
I	intensity
J	classical homogeneous nucleation rate
j	flux
K	kinetic prefactor
k_B	Boltzmann constant
$\vec{k}_{in}$	incoming wave vector
$\vec{k}_{out}$	outgoing scattered wave vector

k_{sp}	equilibrium molecular solubility product
L	length
λ	wavelength
μ	chemical potential
N	number
P	pressure
$P(\vec{q})$	form factor
*pseudo-c***	second boundary concentration in BLG with divalent salts
Φ	volume fraction
$\vec{q}$	momentum transfer
R	ideal gas constant
R_g	radius of gyration
r	radius
r^*	critical cluster size (radius)
ρ	density
S	entropy
$S(\vec{q})$	structure factor
s	relative supersaturation
σ	(hard sphere) diameter
T	temperature
$\mathcal{T}$	transmission
T_c	critical temperature
T_{tr}	transition temperature
t	time
U	internal energy
u	internal energy per unit mass or volume
V	volume
$\tilde{V}$	molar volume

BLG	β-lactoglobulin
CD	circular dichroism
$CdCl_2$	cadmium chloride
$CeCl_3$	cerium chloride
CNT	classical nucleation theory
DLS	dynamic light scattering
FTIR	Fourier transform infrared
$GdCl_3$	gadolinium chloride
$LaCl_3$	lanthanum chloride
LLPS	liquid–liquid phase separation
MIP	metastable intermediate phase
NMR	nuclear magnetic resonance

PDB	protein data bank
pI	isoelectric point
RC	reentrant condensation
SANS	small angle neutron scattering
SAXS	small angle X-ray scattering
SEC	size exclusion chromatography
SLS	static light scattering
UV-Vis	ultraviolet-visible
VSANS	very small angle neutron scattering
$YbCl_3$	ytterbium chloride
$Yb(NO_3)_3$	ytterbium nitrate
$Yb_2(SO_4)_3$	ytterbium sulfate
YCl_3	yttrium chloride
$Y(NO_3)_3$	yttrium nitrate
$Y_2(SO_4)_3$	yttrium sulfate
$ZnCl_2$	zinc chloride

Bibliography

[1] *Protein data bank*, http://www.rcsb.org/pdb/home/home.do, February 2015.

[2] J. J. Abel, *Crystalline insulin*, Proc. Natl. Acad. Sci. USA **12** (1926), 132–136.

[3] Jozef Adamcik, Jin-Mi Jung, Jérôme Flakowski, Paolo De Los Rios, Giovanni Dietler, and Raffaele Mezzenga, *Understanding amyloid aggregation by statistical analysis of atomic force microscopy images*, Nature Nanotechnology **5** (2010), 423–428.

[4] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, *PHENIX: a comprehensive Python-based system for macromolecular structure solution*, Acta Crystallogr. D Biol. Crystallogr. **66** (2010), 213–221.

[5] Ryo Akiyama and Ryo Sakata, *An integral equation study of reentrant behavior in attractive interactions between like-charged macroions immersed in an electrolyte solution*, J. Phys. Soc. Jpn. **80** (2011), no. 12, 123602.

[6] Jens Als-Nielsen and Des McMorrow, *Elements of modern X-ray physics*, John Wiley & Sons, West Sussex, United Kingdom, 2011.

[7] Valerie J. Anderson and Henk N. W. Lekkerkerker, *Insights into phase transition kinetics from colloid science*, Nature **416** (2002), no. 6883, 811–815.

[8] A.K. Arora and B.V.R. Tata, *Ordering and phase transitions in charged colloids*, Complex fluids and fluid microstructures, VCH Cambridge, 1996.

[9] N.W. Ashcroft and J. Lekner, *Structure and resistivity of liquid metals*, Phys. Rev. **145** (1966), 83–90.

[10] Neer Ruben Asherie, *Protein crystallization and phase diagrams*, Methods **34** (2004), no. 3, 266–272.

[11] Stefan Auer and Daan Frenkel, *Prediction of absolute crystal-nucleation rate in hard-sphere colloids*, Nature **409** (2001), 1020–1023.

[12] Young Chan Bae, S. M. Lambert, D. S. Soane, and J. M. Prausnitz, *Cloud-Point Curves of Polymer Solutions from Thermooptical Measurements*, Macromolecules **24** (1991), 4403–4407.

[13] Jillian F. Banfield, Susan A. Welch, Hengzhong Zhang, Tamara Thomsen Ebert, and R. Lee Penn, *Aggregation-based crystal growth and microstructure development in natural iron oxyhydroxide biomineralization products*, Science **289** (2000), no. 5480, 751–754.

[14] Andreas Barth and Christian Zscherp, *What vibrations tell about proteins*, Quarterly Reviews of Biophysics **35** (2002), no. 4, 369–430.

[15] R. J. Baxter, *Percus–Yevick equation for hard spheres with surface adhesion*, J. Chem. Phys. **49** (1968), no. 6, 2770–2774.

[16] G. Beaucage, *Approximations leading to a unified exponential/power-law approach to small-angle scattering*, Journal of Applied Crystallography **28** (1995), no. 6, 717–728.

[17] ———, *Small-angle scattering from polymeric mass fractals of arbitrary mass-fractal dimension*, J. Appl. Cryst. **29** (1996), no. 2, 134–146.

[18] R. Becker and W. Döring, *Kinetische Behandlung der Keimbildung in Übersättigten Dämpfern.*, Ann. Phys. **24** (1935), 719–752.

[19] J. D. Bernal and D. Crowfoot, *X-ray photographs of crystalline pepsin*, Nature **133** (1934), 794–795.

[20] Aleksandar Bijelic, Christian Molitor, Stephan G. Mauracher, Rami Al-Oweini, Ulrich Kortz, and Annette Rompel, *Hen egg-white lysozyme crystallisation: Protein stacking and structure stability enhanced by a tellurium(vi)-centred polyoxotungstate*, ChemBioChem **16** (2015), no. 2, 233–241.

[21] F. Bonneté, S. Finet, and A. Tardieu, *Second virial coefficient: variations with lysozyme crystallization conditions*, J. Cryst. Growth **196** (1999), no. 2-4, 403–414.

[22] F. Bonneté, M. Malfois, S. Finet, A. Tardieu, S. Lafont, and S. Veesler, *Different tools to study interaction potentials in gamma-crystallin solutions: relevance to crystal growth.*, Acta crystallographica. Section D, Biological crystallography **53** (1997), 438–447.

[23] F. Bonneté and D. Vivarès, *Interest of the normalized second virial coefficient and interaction potentials for crystallizing large macromolecules*, Acta Crystallogr., Sect. D **58** (2002), no. 10, 1571–1575.

[24] F. Bonneté, D Vivarès, and C Robert, *Interactions in solution and crystallization of aspergillus flavus urate oxidase*, Journal of crystal growth **232** (2001), 330–339.

[25] Michal Braun, Olga Matsarskaia, Marcell Wolf, Fajun Zhang, Felix Roosen-Runge, Roland Roth, and Frank Schreiber, *Lower critical solution temperature phase behavior in protein solutions induced by multivalent metal ions*, in preparation (2015).

[26] M L Broide, C R Berland, J Pande, O O Ogun, and G B Benedek, *Binary-liquid phase separation of lens protein solutions*, Proc. Natl. Acad. Sci. USA **88** (1991), no. 13, 5660–5664.

[27] S. Brownlow, J. H. Morais Cabral, R. Cooper, D. R. Flower, S. J. Yewdall, I. Polikarpov, A. C. T. North, and L. Sawyer, *Bovine β-lactoglobulin at 1.8 Å resolution - still an enigmatic lipocalin*, Structure **5** (1997), no. 4, 481–495.

[28] Anna Bujacz, *Structures of bovine, equine and leporine serum albumin*, Acta Crystallogr., Sect. D **68** (2012), no. 10, 1278–1289.

[29] Naomi E. Chayen and Emmanuel Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal.*, Nature Methods **5** (2008), no. 2, 147–153.

[30] Qiuying Chen, Peter G. Vekilov, Ronald L. Nagel, and Rhoda Elison Hirsch, *Liquid-liquid phase separation in hemoglobins: Distinct aggregation mechanisms of the β6 mutants*, Biophys. J. **86** (2004), no. 3, 1702–1712.

[31] Sow-Hsin Chen, *Small angle neutron scattering studies of the structure and interaction in micellar and microemulsion systems*, Annu. Rev. Phys. Chem. **37** (1986), no. 1, 351–399.

[32] V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, *Molprobity: all-atom structure validation for macromolecular crystallography*, Acta Crystallogr. D Biol. Crystallogr. **66** (2010), 12–21.

[33] Stephen Z. D. Cheng, *Phase transitions in polymers: the role of metastable states*, 1st ed., Elsevier North-Holland, 2008.

[34] Helmut Cölfen and Markus Antonietti, *Mesocrystals and nonclassical crystallization*, Wiley West Sussex, 2008.

[35] Neil G. Connelly, Ture Damhus, Richard M. Hartshorn, and Alan T. Hutton, *Nomenclature of inorganic chemistry — IUPAC recommendations 2005*, vol. 27, International Union of Pure and Applied Chemistry, Cambridge, January 2005.

[36] Nicola S. A. Davies and Robert D. Gillard, *The solubility loop of nicotine:water*, Transition Metal Chemistry **25** (2000), 628–629.

[37] H.T. Davis, *Statistical mechanics of phases, interfaces, and thin films*, Advances in interfacial engineering series, VCH New York, 1996.

[38] M. Delaye, J.I. Clark, and G.B. Benedek, *Coexistence curves for phase separation in the calf lens cytoplasm*, Biochem. Biophys. Res. Commun. **100** (1981), no. 2, 908–914.

[39] A.V. Delgado, F. González-Caballero, R.J. Hunter, L.K. Koopal, and J. Lyklema, *Measurement and interpretation of electrokinetic phenomena*, J. Colloid Interface Sci. **309** (2007), no. 2, 194–224.

[40] T. Dingermann, D. Steinhilber, G. Folkers, R. Mannhold, and H. Kubinyi, *Molecular biology in medicinal chemistry*, Wiley Weinheim, 2004.

[41] Jonathan P K Doye, Ard A Louis, and Michelle Vendruscolo, *Inhibition of protein crystallization by evolutionary negative design*, Phys. Biol. **1** (2004), 9–13.

[42] Jonathan P K Doye and Wilson C K Poon, *Protein crystallization in vivo*, Curr. Opin. Colloid Interface Sci. **11** (2006), no. 1, 40–46.

[43] S. D. Durbin and G. Feher, *Protein crystallization*, Annu. Rev. Phys. Chem. **47** (1996), no. 1, 171–204.

[44] Ulla M. Elofsson, Marie A. Paulsson, and Thomas Arnebrant, *Adsorption of β-lactoglobulin A and B in relation to self-association:effect of concentration and pH*, Langmuir **13** (1997), no. 6, 1695–1700.

[45] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan, *Features and development of Coot*, Acta Crystallogr. D Biol. Crystallogr. **66** (2010), 486–501.

[46] Eppendorf, *Produkte und Applikationen für das Labor*, 2007-2008, Data Sheet for Eppendorf Pipettes.

[47] Deniz Erdemir, Alfred Y. Lee, and Allan S. Myerson, *Nucleation of crystals from solution: Classical and two-step models*, Acc. Chem. Res. **42** (2009), no. 5, 621–629.

[48] S. Finet, F. Bonneté, J. Frouin, K. Provost, and A. Tardieu, *Lysozyme crystal growth, as observed by small angle X-ray scattering, proceeds without crystallization intermediates.*, European Biophysics Journal **27** (1998), no. 3, 263–271.

[49] D R Flower, *The lipocalin protein family: structure and function*, The Biochemical journal **318** (1996), 1–14.

[50] Gerhard Furrer, Brian L. Phillips, Kai-Uwe Ulrich, Rosemarie Pöthig, and William H. Casey, *The origin of aluminum flocs in polluted streams*, Science **297** (2002), no. 5590, 2245–2247.

[51] Diana Fusco, J. J. Headd, A De Simone, Jun Wang, and Patrick Charbonneau, *Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study*, Soft Matter **10** (2014), 290–302.

[52] Oleg Galkin, Kai Chen, Ronald L. Nagel, Rhoda Elison Hirsch, and Peter G. Vekilov, *Liquid–liquid separation in solutions of normal and sickle cell hemoglobin*, Proc. Natl. Acad. Sci. USA **99** (2002), no. 13, 8479–8483.

[53] Oleg Galkin and Peter G Vekilov, *Are nucleation kinetics of protein crystals similar to those of liquid droplets?*, J. Am. Chem. Soc. **122** (2000), no. 1, 156–163.

[54] _____, *Control of protein crystal nucleation around the metastable liquid–liquid phase boundary*, Proc. Natl. Acad. Sci. USA **97** (2000), no. 12, 6277–6281.

[55] U Gasser, *Crystallization in three- and two-dimensional colloidal suspensions*, Journal of Physics: Condensed Matter **21** (2009), no. 20, 203101.

[56] Denis Gebauer and Helmut Cölfen, *Prenucleation clusters and non-classical nucleation*, Nano Today **6** (2011), no. 6, 564–584.

[57] Denis Gebauer, Matthias Kellermeier, Julian D Gale, Lennart Bergström, and Helmut Cölfen, *Pre-nucleation clusters as solute precursors in crystallisation*, Chemical Society reviews **43** (2014), no. 7, 2348–2371.

[58] Denis Gebauer, Antje Völkel, and Helmut Cölfen, *Stable prenucleation calcium carbonate clusters*, Science **322** (2008), no. 5909, 1819–1822.

[59] A. George and W. W. Wilson, *Predicting protein crystallization from a dilute solution property*, Acta Crystallogr., Sect. D **50** (1994), no. 4, 361–365.

[60] J. W. Gibbs, *On the equilibrium of heterogeneous substances*, Trans. Connect. Acad. Sci. **16** (1876), 343–524.

[61] _____, *On the equilibrium of heterogeneous substances*, Trans. Connect. Acad. Sci. **3** (1878), 108–248.

[62] Richard Giegé, *A historical perspective on protein crystallization from 1840 to the present day*, The FEBS journal **280** (2013), no. 24, 6456–6497.

[63] O. Glatter and O. Kratky, *Small angle X-ray scattering*, Academic Press London, United States Edition New York, 1982.

[64] Olga Gliko, Nikolaus Neumaier, Weichun Pan, Ilka Haase, Markus Fischer, Adelbert Bacher, Sevil Weinkauf, and Peter G. Vekilov, *A metastable prerequisite for the growth of lumazine synthase crystals*, J. Am. Chem. Soc. **127** (2005), no. 10, 3433–3438.

[65] Olga Gliko, Weichun Pan, Panagiotis Katsonis, Nikolaus Neumaier, Oleg Galkin, Sevil Weinkauf, and Peter G. Vekilov, *Metastable liquid clusters in super- and undersaturated protein solutions*, J. Phys. Chem. B **111** (2007), no. 12, 3106–3114.

[66] Christoph Gögelein, Gerhard Nägele, Remco Tuinier, Thomas Gibaud, Anna Stradner, and Peter Schurtenberger, *A simple patchy colloid model for the phase behavior of lysozyme dispersions*, J. Chem. Phys. **129** (2008), no. 8, 085102.

[67] L Gránásy and Gi Tóth, *Colloidal suspense*, Nature Physics **10** (2014), 8–10.

[68] D. W. Green and R. Aschaffenburg, *Twofold symmetry of the β-lactoglobulin molecule in crystals*, J. Mol. Biol. **1** (1959), no. 1, 54–64.

[69] P.R. Griffiths and J.A. De Haseth, *Fourier transform infrared spectrometry*, Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications, Wiley New Jersey, 2007.

[70] J.D. Gunton, A. Shiryayev, and D.L. Pagan, *Protein condensation: Kinetic pathways to crystallization and disease*, Cambridge University Press, 2007.

[71] M. H. J. Hagen and D. Frenkel, *Determination of phase diagrams for the hard-core attractive Yukawa system*, J. Chem. Phys. **101** (1994), no. 5, 4093–4097.

[72] Boualem Hammouda, *A new Guinier-Porod model*, J. Appl. Cryst. **43** (2010), 1474–1478.

[73] Jean-Pierre Hansen and John B. Hayter, *A rescaled MSA structure factor for dilute charged colloidal dispersions*, Mol. Phys. **46** (1982), no. 3, 651–656.

[74] Jean-Pierre Hansen and I. R. McDonald, *Theory of simple liquids*, 3rd ed., Academic Press Amsterdam, 2006.

[75] J. L. Harland, S. I. Henderson, S. M. Underwood, and W. van Megen, *Observation of accelerated nucleation in dense colloidal fluids of hard sphere particles*, Phys. Rev. Lett. **75** (1995), 3572–3575.

[76] Thomas K. Haxton and Stephen Whitelam, *Design rules for the self-assembly of a protein crystal*, Soft Matter **8** (2012), 3558–3562.

[77] John B. Hayter and Jeff Penfold, *An analytic structure factor for macroion solutions*, Mol. Phys. **42** (1981), no. 1, 109–118.

[78] Franz Hofmeister, *Zur Lehre von der Wirkung der Salze*, N-S Arch. Pharmacol. **24** (1888), 247–260.

[79] Shelby B. Hutchens and Zhen-Gang Wang, *Metastable cluster intermediates in the condensation of charged macromolecule solutions*, J. Chem. Phys. **127** (2007), no. 8, 084912.

[80] Luca Ianeselli, Fajun Zhang, Maximilian W A Skoda, Robert M J Jacobs, Richard A Martin, Shirley Callow, Sylvain Prévost, and Frank Schreiber, *Protein-protein interactions in ovalbumin solutions studied by small-angle scattering: effect of ionic strength and the chemical nature of cations*, J. Phys. Chem. B **114** (2010), no. 11, 3776–3783.

[81] Coe Ishimoto and Toyoichi Tanaka, *Critical behavior of a binary mixture of protein and salt water*, Phys. Rev. Lett. **39** (1977), no. 8, 474–477.

[82] A. Isihara, *Determination of molecular shape by osmotic measurement*, J. Chem. Phys. **18** (1950), no. 11, 1446–1449.

[83] Liam C Jacobson, Waldemar Hujo, and Valeria Molinero, *Amorphous precursors in the nucleation of clathrate hydrates*, J. Am. Chem. Soc. **132** (2010), no. 33, 11806–11811.

[84] Shweta Jain and Jayant B Udgaonkar, *Salt-induced modulation of the pathway of amyloid fibril formation by the mouse prion protein.*, Biochemistry **49** (2010), no. 35, 7615–7624.

[85] Richard A. L. Jones, *Soft condensed matter*, Oxford University Press, USA, 2002.

[86] Elena Jordan, Felix Roosen-Runge, Sara Leibfarth, Fajun Zhang, Michael Sztucki, Andreas Hildebrandt, Oliver Kohlbacher, and Frank Schreiber, *Competing salt effects on phase behavior of protein solutions: Tailoring of protein interaction by the binding of multivalent ions and charge screening*, J. Phys. Chem. B **118** (2014), no. 38, 11365–11374.

[87] Wolfgang Kabsch, *XDS*, Acta Crystallographica Section D **66** (2010), no. 2, 125–132.

[88] Takeshi Kawasaki and Hajime Tanaka, *Formation of a crystal nucleus from liquid*, Proceedings of the National Academy of Sciences **107** (2010), no. 32, 14036–14041.

[89] Sharon M. Kelly, Thomas J. Jess, and Nicholas C. Price, *How to study proteins by circular dichroism*, Biochimica et Biophysica Acta **1751** (2005), no. 2, 119–139.

[90] Steven R. Kline, *Reduction and analysis of SANS and USANS data using IGOR Pro*, J. Appl. Cryst. **39** (2006), no. 6, 895–900.

[91] E. Krissinel and K. Henrick, *Inference of macromolecular assemblies from crystalline state*, J. Mol. Biol. **372** (2007), no. 3, 774–797.

[92] Anna Kubíčková, Tomáš Křížek, Pavel Coufal, Mario Vazdar, Erik Wernersson, Jan Heyda, and Pavel Jungwirth, *Overcharging in biological systems: reversal of electrophoretic mobility of aqueous polyaspartate by multivalent cations*, Phys. Rev. Lett. **108** (2012), 186101.

[93] Amit M. Kulkarni, Narendra M. Dixit, and Charles F. Zukoski, *Ergodic and non-ergodic phase transitions in globular protein suspensions*, Faraday Discuss. **123** (2003), 37–50.

[94] Yu Kuznetsov, A Malkin, and A McPherson, *Atomic-force-microscopy studies of phase separations in macromolecular systems*, Phys. Rev. B **58** (1998), no. 10, 6097–6103.

[95] Albert L. Lehninger, David L. Nelson, and Michael M. Cox, *Lehninger principles of biochemistry*, W. H. Freeman New York, 2005.

[96] Ye Li, Vassiliy Lubchenko, and Peter G. Vekilov, *The use of dynamic light scattering and Brownian microscopy to characterize protein aggregation*, The Review of scientific instruments **82** (2011), no. 5, 053106.

[97] Ye Li, Vassiliy Lubchenko, Maria A. Vorontsova, Luis Filobelo, and Peter G Vekilov, *Ostwald-like ripening of the anomalous mesoscopic clusters in protein solutions*, J. Phys. Chem. B **116** (2012), no. 35, 10657–10664.

[98] B. Linder, *Thermodynamics and introductory statistical mechanics*, Wiley InterScience online books, Wiley New Jersey, 2004.

[99] P. Lindner and T. Zemb, *Neutrons, X-rays, and light: Scattering methods applied to soft condensed matter*, Elsevier North-Holland, 2002.

[100] Yingxin Liu, Xiujuan Wang, and Chi Bun Ching, *Toward further understanding of lysozyme crystallization: Phase diagram, protein-protein interaction, nucleation kinetics, and growth kinetics*, Cryst. Growth Des. **10** (2010), no. 2, 548–558.

[101] Yun Liu, Emiliano Fratini, Piero Baglioni, Wei-Ren Chen, and Sow-Hsin Chen, *Effective long-range attraction between protein molecules in solutions studied by small angle neutron scattering*, Phys. Rev. Lett. **95** (2005), 118102.

[102] Aleksey Lomakin, Neer Ruben Asherie, and George B. Benedek, *Liquid-solid transition in nuclei of protein crystals*, PNAS **100** (2003), no. 18, 10254–10257.

[103] A. W. Loven and O. K. Rice, *Coexistence curve of the 2,6-lutidine+water system in the critical region*, Transactions of the Faraday Society **59** (1963), 2723–2727.

[104] Mikael Lund, Pavel Jungwirth, and C.E. Woodward, *Ion specific protein assembly and hydrophobic surface forces*, Phys. Rev. Lett. **100** (2008), no. 25, 258105.

[105] Mikael Lund, Robert Vacha, and Pavel Jungwirth, *Specific ion binding to macromolecules: effects of hydrophobicity and ion pairing.*, Langmuir **24** (2008), no. 7, 3387–3391.

[106] James F Lutsko, *Nucleation of colloids and macromolecules: Does the nucleation pathway matter?*, J. Chem. Phys. **136** (2012), 134502.

[107] James F. Lutsko and Grégoire Nicolis, *Theoretical evidence for a dense fluid precursor to crystallization*, Phys. Rev. Lett. **96** (2006), 046102.

[108] Johannes Lyklema, *Overcharging, charge reversal: Chemistry or physics?*, Colloid Surface A **291** (2006), no. 1-3, 3–12.

[109] Karolina A. Majorek, Przemyslaw J. Porebski, Arjun Dayal, Matthew D. Zimmerman, Kamila Jablonska, Alan J. Stewart, Maksymilian Chruszcz, and Wladek Minor, *Structural and immunologic characterization of bovine, horse, and rabbit serum albumins*, Mol. Immunol. **52** (2012), no. 3–4, 174–182.

[110] G. N. Malcolm and J. S. Rowlinson, *The thermodynamic properties of aqueous solutions of polyethylene glycol, polypropylene glycol and dioxane*, Transactions of the Faraday Society **53** (1957), 921–931.

[111] A. J. Malkin, Y. G. Kuznetsov, T.A. Land, J. J. DeYoreo, and A. McPherson, *Mechanisms of growth for protein and virus crystals*, Nature Structural Biology **2** (1995), 956–959.

[112] Malvern Instruments, *Zetasizer nano user manual*, 2008.

[113] S. Martin, G. Bryant, and W. van Megen, *Crystallization kinetics of polydisperse colloidal hard spheres: Experimental evidence for local fractionation*, Phys. Rev. E **67** (2003), 061405.

[114] ———, *Crystallization kinetics of polydisperse colloidal hard spheres. II. Binary mixtures*, Physical Review E **71** (2005), 021404.

[115] Olga Matsarskaia, *Influence of trivalent metal ions on phase behaviour in protein solutions*, Master's thesis, Eberhard Karls University Tuebingen, 2014.

[116] Stephan Gerhard Mauracher, Christian Molitor, Rami Al-Oweini, Ulrich Kortz, and Annette Rompel, *Crystallization and preliminary X-ray crystallographic analysis of latent isoform PPO4 mushroom (Agaricus bisporus) tyrosinase*, Acta crystallographica. Section F, Structural biology communications **70** (2014), no. Pt 2, 263–266.

[117] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, *Phaser crystallographic software*, J. Appl. Crystallogr. **40** (2007), 658–674.

[118] T. L. McMeekin, M. L. Groves, and N. J. Hipp, *Partial specific volume of the protein and water in beta-lactoglobulin crystals*, Journal of Polymer Science **XII** (1954), 309–315.

[119] A. McPherson, A. J. Malkin, and Y. Kuznetsov, *Atomic force microscopy in the study of macromolecular crystal growth*, Annu. Rev. Biophys. Biomol. Struct. **29** (2000), 361–410.

[120] Alexander McPherson, *Current approaches to macromolecular crystallization*, European Journal of Biochemistry **189** (1990), no. 1, 1–23.

[121] Allen P. Minton, *The effect of time-dependent macromolecular crowding on the kinetics of protein aggregation: a simple model for the onset of age-related neurodegenerative disease*, Frontiers in Physics **2** (2014), 1–5.

[122] Svetlana Mintova, Norman H. Olson, Valentin Valtchev, and Thomas Bein, *Mechanism of zeolite a nanocrystal growth from colloids at room temperature*, Science **283** (1999), no. 5404, 958–960.

[123] Christian Moitzi, Laurence Donato, Christophe Schmitt, Lionel Bovetto, Graeme Gillies, and Anna Stradner, *Structure of β-lactoglobulin microgels formed during heating as revealed by small-angle X-ray scattering and light scattering*, Food Hydrocolloids **25** (2011), no. 7, 1766–1774.

[124] Vicki Mountain, *Astex, Structural Genomix, and Syrrx - I can see clearly now: structural biology and drug discovery*, Chemistry & Biology **10** (2003), 95–98.

[125] G. N. Murshudov, A. A. Vagin, and E. J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method*, Acta Crystallogr. D Biol. Crystallogr. **53** (1997), 240–255.

[126] Martin Muschol and Franz Rosenberger, *Liquid–liquid phase separation in supersaturated lysozyme solutions and associated precipitate formation/crystallization*, J. Chem. Phys. **107** (1997), no. 6, 1953–1962.

[127] Christo N. Nanev, *Kinetics and intimate mechanism of protein crystal nucleation*, Progress in Crystal Growth and Characterization of Materials **59** (2013), no. 4, 133–169.

[128] Janaky Narayanan and X Y Liu, *Protein interactions in undersaturated and supersaturated solutions: a study using light and X-ray scattering*, Biophys. J. **84** (2003), 523–532.

[129] T. Narayanan, *Soft matter characterization*, ch. 17: Synchroton small-angle X-ray Scattering, Springer New York, 2008.

[130] T. Narayanan and Anil Kumar, *Reentrant phase transitions in multicomponent liquid mixtures*, Physics Reports **249** (1994), 135–218.

[131] Alexandra Navrotsky, *Energetic clues to pathways to biomineralization: precursors, clusters, and nanoparticles*, Proc. Natl. Acad. Sci. USA **101** (2004), no. 33, 12096–12101.

[132] Marcia E. Newcomer, Anders Liljas, Ulf Eriksson, Johan Sundelin, Lars Rask, and Per A. Peterson, *Crystallization of and preliminary X-ray data for an intracellular vitamin A-binding protein from rat liver*, Journal of Biological Chemistry **256** (1981), no. 15, 8162–8163.

[133] T. T. Nguyen, I. Rouzina, and B. I. Shklovskii, *Reentrant condensation of DNA induced by multivalent counterions*, J. Chem. Phys. **112** (2000), no. 5, 2562–2568.

[134] T. T. Nguyen and B. I. Shklovskii, *Complexation of DNA with positive spheres: Phase diagram of charge inversion and reentrant condensation*, J. Chem. Phys. **115** (2001), no. 15, 7298–7308.

[135] T Nicolai, M Pouzot, D Durand, M Weijers, and R. W Visschers, *Isoscattering points during heat-induced aggregation and gelation of globular proteins indicating micro-phase separation*, Europhysics Letters (EPL) **73** (2006), no. 2, 299–305.

[136] W. Ostwald, *Studien über die Bildung und Umwandlung fester Körper*, Zeitschrift für physikalische Chemie **22** (1897), 289–330.

[137] Weichun Pan, Oleg Galkin, Luis Filobelo, Ronald L. Nagel, and Peter G. Vekilov, *Metastable mesoscopic clusters in solutions of sickle-cell hemoglobin*, Biophys. J. **92** (2007), no. 1, 267–277.

[138] Weichun Pan, Peter G Vekilov, and Vassiliy Lubchenko, *Origin of anomalous mesoscopic phases in protein solutions.*, J. Phys. Chem. B **114** (2010), no. 22, 7620–7630.

[139] P. Papon, S.L. Schnur, J. Leblond, and P.H.E. Meijer, *The physics of phase transitions: Concepts and applications*, Advanced Texts in Physics, Springer Berlin Heidelberg, 2006.

[140] A. K Patra, S. Ramanathan, D. Sen, and S. Mazumder, *SANS study of fractal microstructure and pore morphology in porous titania*, Journal of Alloys and Compounds **397** (2005), 300–305.

[141] Yi Peng, Feng Wang, Ziren Wang, Ahmed M. Alsayed, Zexin Zhang, Arjun G. Yodh, and Yilong Han, *Two-step nucleation mechanism in solid–solid phase transitions*, Nat. Mater. **14** (2015), 101–108.

[142] Roberto Piazza, *Interactions and phase transitions in protein solutions*, Curr. Opin. Colloid Interface Sci. **5** (2000), no. 1-2, 38–43.

[143] ———, *Protein interactions and association: an open challenge for colloid science*, Curr. Opin. Colloid Interface Sci. **8** (2004), no. 6, 515–522.

[144] V. Pipich, *QtiKWS: user-friendly program for reduction, visualization, analysis and fit of SA(N)S data*, http://www.qtikws.de, 2012.

[145] Christopher D. Putnam, Michal Hammel, Greg L. Hura, and John A. Tainer, *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution*, Quarterly Reviews of Biophysics **40** (2007), no. 3, 191–285.

[146] Bin Y. Qin, Geoffrey B. Jameson, Maria C. Bewley, Edward N. Baker, and Lawrence K. Creamer, *Functional implications of structural differences between variants A and B of bovine β-lactoglobulin*, Protein Science **8** (1999), no. 1, 75–83.

[147] L Ratke and PW Voorhees, *Growth and coarsening: Ostwald ripening in material processing*, Springer Berlin Heidelberg, 2002.

[148] William J Ray and CE Bracker, *Polyethylene glycol: catalytic effect on the crystallization of phosphoglucomutase at high salt concentration*, Journal of Crystal Growth **76** (1986), 562–576.

[149] P. Richet, *The physical basis of thermodynamics: With applications to chemistry*, Springer New York, 2001.

[150] E. Riedel and C. Janiak, *Anorganische Chemie*, de Gruyter Berlin, 2007.

[151] Felix Roosen-Runge, Benjamin S. Heck, Fajun Zhang, Oliver Kohlbacher, and Frank Schreiber, *Interplay of pH and binding of multivalent metal ions: Charge inversion and reentrant condensation in protein solutions*, J. Phys. Chem. B **117** (2013), no. 18, 5777–5787.

[152] Felix Roosen-Runge, Fajun Zhang, Frank Schreiber, and Roland Roth, *Ion-activated attractive patches as a mechanism for controlled protein interactions*, Sci. Rep. **4** (2014), no. 7016, 7016.

[153] D. Rosenbaum, *Protein interactions and crystallization*, J. Cryst. Growth **169** (1996), no. 169, 752–758.

[154] D Rosenbaum, PC Zamora, and CF Zukoski, *Phase behavior of small attractive colloidal particles*, Phys. Rev. Lett. **76** (1996), no. 1, 150–153.

[155] C.A. Ross and M.A. Poirier, *Protein aggregation and neurodegenerative disease*, Nat. Med. **10** (2004), 10–17.

[156] Takaaki Sato, Togo Shimozawa, Toshiko Fukasawa, Masako Ohtaki, Kenji Aramaki, Katsuzo Wakabayashi, and Shin'ichi Ishiwata, *Actin oligomers at the initial stage of polymerization induced by increasing temperature at low ionic strength: study with small-angle X-ray scattering*, Biophysics **6** (2010), 1–11.

[157] Andrea Sauter, *Crystallization and phase behavior of aqueous beta-lactoglobulin solutions in the presence of multivalent cations*, Diplomarbeit, Eberhard Karls University Tuebingen, 2011.

[158] Andrea Sauter, Melanie Oelker, Georg Zocher, Fajun Zhang, Thilo Stehle, and Frank Schreiber, *Nonclassical pathways of protein crystallization in the presence of multivalent metal ions*, Crystal Growth and Design **14** (2014), no. 12, 6357–6366.

[159] Andrea Sauter, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Artem Feoktystov, Robert M. J. Jacobs, and Frank Schreiber, *On the question of two-step nucleation in protein crystallization*, Faraday Discussions **179** (2015), 41–58.

[160] Andrea Sauter, Felix Roosen-Runge, Fajun Zhang, Gudrun Lotze, Robert M. J. Jacobs, and Frank Schreiber, *Real-time observation of nonclassical protein crystallization kinetics*, J. Am. Chem. Soc. **137** (2015), no. 4, 1485–1491.

[161] J. R. Savage and A. D. Dinsmore, *Experimental evidence for two-step nucleation in colloidal crystallization*, Phys. Rev. Lett. **102** (2009), 198302.

[162] Lindsay Sawyer, George Kontopidis, and Su-Ying Wu, *β-lactoglobulin – a three-dimensional perspective*, International Journal of Food Science and Technology **34** (1999), 409–418.

[163] Howard G. Schild and David A. Tirrell, *Microcalorimetric detection of lower critical solution temperatures in aqueous polymer solutions*, Journal of Physical Chemistry **94** (1990), 4352–4356.

[164] T. Schilling, H. J. Schöpe, M. Oettel, G. Opletal, and I. Snook, *Precursor-mediated crystallization process in suspensions of hard spheres*, Phys. Rev. Lett. **105** (2010), no. 2, 025701.

[165] J. Schlichtkrull, *Insulin crystals*, Acta Chemica Scandinavica **10** (1956), no. 9, 1455–1458.

[166] Christophe Schmitt, Christian Moitzi, Claudine Bovay, Martine Rouvet, Lionel Bovetto, Laurence Donato, Martin E. Leser, Peter Schurtenberger, and Anna Stradner, *Internal structure and colloidal behaviour of covalent whey protein microgels obtained by heat treatment*, Soft Matter **6** (2010), no. 19, 4876–4884.

[167] C.A. Schneider, W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis*, Nature Methods **9** (2012), 671–675.

[168] Hans Joachim Schöpe, Gary Bryant, and William van Megen, *Two-step crystallization kinetics in colloidal hard-sphere systems*, Phys. Rev. Lett. **96** (2006), 175701.

[169] LLC Schrodinger, *The pymol molecular graphics system, version 1.4.1*, The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrodinger, LLC., 2010.

[170] Anja Schuetz, Didier Nana, Charlotte Rose, Georg Zocher, Maja Milanovic, Jessica Koenigsmann, Rosel Blasig, Udo Heinemann, and Dirk Carstanjen, *The structure of the klf4 DNA-binding domain links to self-renewal and macrophage differentiation*, Cellular and Molecular Life Sciences **68** (2011), no. 18, 3121–3131.

[171] R. P. Sear, *The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate*, International Materials Reviews **57** (2012), no. 6, 328–356.

[172] Richard Sear, *Nucleation at contact lines where fluid–fluid interfaces meet solid surfaces*, J. Phys.: Condens. Matter **19** (2007), 466106.

[173] Richard P. Sear, *Interactions in protein solutions*, Curr. Opin. Colloid Interface Sci. **11** (2006), no. 1, 35–39.

[174] Richard P Sear, *Nucleation via an unstable intermediate phase*, The Journal of chemical physics **131** (2009), no. 7, 074702.

[175] K. Seeger, K. Flinspach, E. Haug-Schifferdecker, A. Kulik, B. Gust, H. Fiedler, and L. Heide, *The biosynthetic genes for prenylated phenazines are located at two different chromosomal loci of Streptomyces cinnamonensis DSM 1042*, Microbial Biotechnology **4** (2011), no. 2, 252–262.

[176] R. D. Shannon, *Revised effective ionic radii and systematic studies of interatomic distances in halides and chaleogenides*, Acta Cryst. A **32** (1976), 751–767.

[177] Anuj Shukla, Efstratios Mylonas, Emanuela Di Cola, Stephanie Finet, Peter Timmins, Theyencheri Narayanan, and Dmitri I. Svergun, *Absence of equilibrium cluster phase in concentrated lysozyme solutions*, Proc. Natl. Acad. Sci. USA **105** (2008), no. 13, 5075–5080.

[178] Sigma-Aldrich, *Product information*, 2011.

[179] D.S. Sivia, *Elementary scattering theory: for X-ray and neutron users*, OUP Oxford, 2011.

[180] Mike Sleutel, Jim Lutsko, Alexander E.S. van Driessche, Miguel A. Durán-Olivencia, and Dominique Maes, *Observing classical nucleation theory at work by monitoring phase transitions with molecular precision*, Nature Communications **5** (2014), 1–8.

[181] Mike Sleutel and Alexander E. S. van Driessche, *Role of clusters in nonclassical nucleation and growth of protein crystals.*, Proc. Nat. Sci. USA **111** (2014), no. 5, E546–E553.

[182] Herbert A. Sober, *CRC Handbook of Biochemistry: Selected data for molecular biology*, The Chemical Rubber Co. Cleveland, Ohio, 1970.

[183] Daniel Soraruf, Felix Roosen-Runge, Marco Grimaldo, Fabio Zanini, Ralf Schweins, Tilo Seydel, Fajun Zhang, Roland Roth, Martin Oettel, and Frank Schreiber, *Protein cluster formation in aqueous solution in the presence of multivalent metal ions - a light scattering study*, Soft Matter **10** (2014), 894–902.

[184] Anna Stradner, Helen Sedgwick, Frederic Cardinaux, Wilson C. K. Poon, Stefan U. Egelhaaf, and Peter Schurtenberger, *Equilibrium cluster formation in concentrated protein solutions and colloids*, Nature **432** (2004), 492–495.

[185] Gert Strobl, *The physics of polymers*, third ed., ch. 5, Springer, Berlin Heidelberg, 2007.

[186] D. Svergun, C. Barberato, and M. H. J. Koch, *CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates*, J. Appl. Cryst. **28** (1995), no. 6, 768–773.

[187] Dmitri I Svergun and Michel H J Koch, *Small-angle scattering studies of biological macromolecules in solution*, Rep. Prog. Phys. **66** (2003), no. 10, 1735–1782.

[188] Dmitri I. Svergun, Michel H. J. Koch, Peter A. Timmins, and Roland P. May, *Small angle X-ray and neutron scattering from solutions of biological macromolecules*, Oxford University Press, 2013.

[189] V. Talanquer and David W. Oxtoby, *Crystal nucleation in the presence of a metastable critical point*, J. Chem. Phys. **109** (1998), no. 1, 223–227.

[190] Peng Tan, Ning Xu, and Lei Xu, *Visualizing kinetic pathways of homogeneous nucleation in colloidal crystallization*, Nat. Phys. **10** (2014), no. 1, 73–79.

[191] Hajime Tanaka and Yuya Nishikawa, *Viscoelastic phase separation of protein solutions*, Physical Review Letters **95** (2005), 078103.

[192] Shinpei Tanaka, Mitsuo Ataka, and Kohzo Ito, *Pattern formation and coarsening during metastable phase separation in lysozyme solutions*, Physical Review E **65** (2002), no. 5, 051804.

[193] T Tanaka and G B Benedek, *Observation of protein diffusivity in intact human and bovine lenses with application to cataract*, Invest. Ophth. Vis. Sci. **14** (1975), no. 6, 449–456.

[194] T Tanaka, C Ishimoto, and LT Chylack, *Phase separation of a protein-water mixture in cold cataract in the young rat lens*, Science **197** (1977), no. 4307, 1010–1012.

[195] L. Ten, *Crystallographic fast fourier transforms*, Acta Crystallographica Section A **29** (1973), no. 2, 183–191.

[196] Pieter Rein ten Wolde and Daan Frenkel, *Enhancement of protein crystal nucleation by critical density fluctuations*, Science **277** (1997), no. 5334, 1975–1978.

[197] ———, *Homogeneous nucleation and the Ostwald step rule*, Physical Chemistry Chemical Physics **1** (1999), no. 9, 2191–2196.

[198] Gyula I. Tóth, Tamás Pusztai, György Tegze, Gergely Tóth, and László Gránásy, *Amorphous nucleation precursor in highly nonequilibrium fluids*, Phys. Rev. Lett. **107** (2011), 175702.

[199] Sergei Trakhanov, David I. Kreimer, Sean Parkin, Giovanna Ferro-Luzzi Ames, and Bernhard Rupp, *Cadmium-induced crystallization of proteins: II. Crystallization of the salmonella typhimurium histidine-binding protein in complex with L-histidine, L-arginine, or L-lysine*, Protein Science **7** (1998), no. 3, 600–604.

[200] Sergei Trakhanov and Florante A. Quiocho, *Influence of divalent cations in protein crystallization*, Protein Science **4** (1995), no. 9, 1914–1919.

[201] Alexander E. S. van Driessche, Juan Manuel García-Ruiz, José Manuel Delgado-López, and Gen Sazaki, *In situ observation of step dynamics on gypsum crystals*, Crystal Growth & Design **10** (2010), no. 9, 3909–3916.

[202] Peter G. Vekilov, *Dense liquid precursor for the nucleation of ordered solid phases from solution*, Cryst. Growth Des. **4** (2004), no. 4, 671–685.

[203] ———, *Phase transitions of folded proteins*, Soft Matter **6** (2010), no. 21, 5254–5272.

[204] ———, *The two-step mechanism of nucleation of crystals in solution*, Nanoscale **2** (2010), 2346–2357.

[205] Peter G. Vekilov, Angela R. Feeling-Taylor, Siu-Tung Yau, and Dimiter Petsev, *Solvent entropy contribution to the free energy of protein crystallization*, Acta Crystallogr., Sect. D **58** (2002), no. 10 Part 1, 1611–1616.

[206] O D Velev, E W Kaler, and A M Lenhoff, *Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen.*, Biophys. J. **75** (1998), no. 6, 2682–2697.

[207] M. Verheul, J. S. Pedersen, S. P. F. M. Roefs, and K. G. de Kruif, *Association behavior of native β-lactoglobulin*, Biopolymers **49** (1999), no. 1, 11–20.

[208] Denis Vivarès, Eric W. Kaler, and Abraham M. Lenhoff, *Quantitative imaging by confocal scanning fluorescence microscopy of protein crystallization via liquid–liquid phase separation*, Acta Crystallogr., Sect. D **61** (2005), no. 6, 819–825.

[209] G. A. Vliegenthart and H. N. W. Lekkerkerker, *Predicting the gas–liquid critical point from the second virial coefficient*, J. Chem. Phys. **112** (2000), no. 12, 5364–5369.

[210] M. Volmer and A. Weber, *Germ-formation in oversaturated figures*, Z. Phys. Chem. **119** (1926), 277–301.

[211] Adam F. Wallace, Lester O. Hedges, Alejandro Fernandez-Martinez, Paolo Raiteri, Julian D. Gale, Glenn A. Waychunas, Stephen Whitelam, Jillian F. Banfield, and James J. De Yoreo, *Microscopic evidence for liquid-liquid separation in supersaturated $CaCO_3$ solutions*, Science **341** (2013), 885–889.

[212] Ying Wang, Aleksey Lomakin, Ramil F. Latypov, and George B. Benedek, *Phase separation in solutions of monoclonal antibodies and the effect of human serum albumin*, Proc. Natl. Acad. Sci. USA **108** (2011), no. 40, 16606–16611.

[213] Stephen Whitelam, *Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions*, Phys. Rev. Lett. **105** (2010), no. 8, 088102.

[214] Stephen Whitelam, *Nonclassical assembly pathways of anisotropic particles*, J. Chem. Phys. **132** (2010), no. 19, 194901.

[215] Marcell Wolf, *Effective interactions in liquid–liquid phase separated protein solutions induced by multivalent ions*, Ph.D. thesis, Eberhard Karls University Tuebingen, 2014.

[216] Marcell Wolf, Felix Roosen-Runge, Fajun Zhang, Roland Roth, Maximilian W. A. Skoda, Robert M. J. Jacobs, Michael Sztucki, and Frank Schreiber, *Effective interactions in protein–salt solutions approaching liquid–liquid phase separation*, Journal of Molecular Liquids **200, Part A** (2014), 20–27, Proceedings of the 7th Mini-Symposium on Liquids "Liquid-Liquid Phase Separation and Related Topics on Liquids".

[217] Scott A. Wood, *The aqueous geochemistry of the rare-earth elements and yttrium: 1. Review of available low-temperature data for inorganic complexes and the inorganic REE speciation of natural waters*, Chem. Geol. **82** (1990), 159–186.

[218] Nanhua Yao, Sergei Trakhanov, and Florante A. Quiocho, *Refined 1.89 Å structure of the histidine-binding protein complexed with histidine and its relationship with many other active transport/chemosensory proteins*, Biochemistry **33** (1994), no. 16, 4769–4779.

[219] Giuliano Zanchetta and Roberto Cerbino, *Exploring soft matter with X-rays: from the discovery of the DNA structure to the challenges of free electron lasers*, Journal of Physics: Condensed Matter **22** (2010), no. 32, 323102.

[220] F. Zhang, M. W. A. Skoda, R. M. J. Jacobs, S. Zorn, R. A. Martin, C. M. Martin, G. F. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher, and F. Schreiber, *Reentrant condensation of proteins in solution induced by multivalent counterions*, Phys. Rev. Lett. **101** (2008), no. 14, 148101.

[221] Fajun Zhang, Felix Roosen-Runge, Andrea Sauter, Roland Roth, Maximilian W. A. Skoda, Robert Jacobs, Michael Sztucki, and Frank Schreiber, *The role of cluster formation and metastable liquid-liquid phase separation in protein crystallization*, Faraday Discuss. **159** (2012), 313–325.

[222] Fajun Zhang, Felix Roosen-Runge, Andrea Sauter, Marcell Wolf, Robert M. J. Jacobs, and Frank Schreiber, *Reentrant condensation, liquid-liquid phase separation and crystallization in protein solutions induced by multivalent metal ions*, Pure Appl. Chem. **86** (2014), 191–202.

[223] Fajun Zhang, Felix Roosen-Runge, Maximilian W. A. Skoda, Robert M. J. Jacobs, Marcell Wolf, Philip Callow, Henrich Frielinghaus, Vitaliy Pipich, Sylvain Prevost, and Frank Schreiber, *Hydration and interactions in protein solutions containing concentrated electrolytes studied by small-angle scattering*, Phys. Chem. Chem. Phys. **14** (2012), 2483–2493.

[224] Fajun Zhang, Roland Roth, Marcell Wolf, Felix Roosen-Runge, Maximilian W. A. Skoda, Robert M. J. Jacobs, Michael Sztucki, and Frank Schreiber, *Charge-controlled metastable liquid-liquid phase separation in protein solutions as a universal pathway towards crystallization*, Soft Matter **8** (2012), 1313–1316.

[225] Fajun Zhang, Maximilian W. A. Skoda, Robert M. J. Jacobs, Richard A. Martin, Christopher M. Martin, and Frank Schreiber, *Protein interactions studied by SAXS: Effect of ionic strength and protein concentration for BSA in aqueous solutions*, J. Phys. Chem. B **111** (2007), no. 1, 251–259.

[226] Fajun Zhang, Sophie Weggler, Michael J. Ziller, Luca Ianeselli, Benjamin S. Heck, Andreas Hildebrandt, Oliver Kohlbacher, Maximilian W. A. Skoda, Robert M. J. Jacobs, and Frank Schreiber, *Universality of protein reentrant condensation in solution induced by multivalent metal ions*, Proteins **78** (2010), no. 16, 3450–3457.

[227] Fajun Zhang, Georg Zocher, Andrea Sauter, Thilo Stehle, and Frank Schreiber, *Novel approach to controlled protein crystallization through ligandation of yttrium cations*, J. Appl. Cryst. **44** (2011), no. 4, 755–762.

[228] Tian Hui Zhang and Xiang Yang Liu, *How does a transient amorphous precursor template crystallization*, J. Am. Chem. Soc. **129** (2007), no. 44, 13520–13526.

[229] Georg Zocher, Orwah Saleh, Joel B. Heim, Dominik A. Herbst, Lutz Heide, and Thilo Stehle, *Structure-based engineering increased the catalytic turnover rate of a novel phenazine prenyltransferase*, PLoS One **7** (2012), no. 10, e48427.

List of Own Publications

The publications this thesis is based on are marked with **[A]**-**[D]**.

- F. Zhang, G. Zocher, A. Sauter, T. Stehle, and F. Schreiber.
  Novel approach to controlled protein crystallization through ligandation of yttrium cations.
  J. Appl. Cryst. **44** (2011), 755–762.

- F. Zhang, F. Roosen-Runge, A. Sauter, R. Roth, M. W. A. Skoda, R. M. J. Jacobs, M. Sztucki and F. Schreiber.
  The role of cluster formation and metastable liquid-liquid phase separation in protein crystallization.
  Faraday Discuss. **159** (2012), 313–325.

- F. Zhang, F. Roosen-Runge, A. Sauter, M. Wolf, R. M. J. Jacobs, and F. Schreiber.
  Reentrant condensation, liquid-liquid phase separation and crystallization in protein solutions induced by multivalent metal ions.
  Pure Appl. Chem. **86** (2014), 191–202.

- **[A]** A. Sauter, M. Oelker, G. Zocher, F. Zhang, T. Stehle, and F. Schreiber.
  Nonclassical pathways of protein crystallization in the presence of multivalent metal ions.
  Cryst. Growth Des. **14** (2014), 6357–6366.

- **[B]** A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, R. M. J. Jacobs, and F. Schreiber.
  Real-time observation of nonclassical protein crystallization kinetics.
  J. Am. Chem. Soc. **137** (2015), 1485–1491.

- **[C]** A. Sauter, F. Roosen-Runge, F. Zhang, G. Lotze, A. Feoktystov, R. M. J. Jacobs, and F. Schreiber.
  On the question of two-step nucleation in protein crystallization.
  Faraday Discuss. (2015), in print.

- **[D]** A. Sauter, F. Zhang, N. K. Szekely, V. Pipich, M. Sztucki and F. Schreiber.
  Structural evolution of metastable protein aggregates in the presence of trivalent salt studied by (V)SANS and SAXS.
  in preparation.

Acknowledgment

The scientific work presented in this thesis was only made possible through the support, collaboration, and contributions of numerous people which I would like to acknowledge in the following.

First of all, I would like to thank my Ph.D. supervisor Prof. Dr. Frank Schreiber for giving me the opportunity of this work and for his good advice.

I thank PD Dr. Fajun Zhang for the excellent mentoring, for all his good ideas, help and support. I enjoyed working with him and learned so many things.

Moreover, I want to thank Prof. Dr. Thilo Stehle for agreeing to be my second referee as well as for the collaboration and the possibility to use his lab resources. Likewise, I thank Georg Zocher and Melanie Oelker for the productive collaboration during the last years.

I benefited from the fruitful discussions with Dr. Felix Roosen-Runge, Prof. Ph.D. Robert Jacobs and many others and want to thank them for the cooperation.

"P19 and friends" deserve special thanks for excellent team work and a lot of fun. I also gratefully acknowledge the very helpful proofreading of my thesis by Olga Matsarskaia, Michal Braun, Stefano Da Vela, Marcell Wolf, Marco Grimaldo, Christoph Wolber, Jonathan Koschera, Fajun Zhang, and Frank Schreiber. All group members and especially the coworkers from the protein subgroup deserve my thanks for the common research, all the productive discussions and the good team work during beamtimes.

I thank the ESRF and the FRM2 for the allocation of beamtime and the beamline staffs for their support, in particular the beamline scientists Gudrun Lotze, Artem Feoktystov, Michael Sztucki, Noemi K. Szekely, and Vitaliy Pipich.

I acknowledge the Landesgraduiertenförderung for financial support. I am grateful to Aleksandra Rötschke and Hanna Maurer for their manifold help in bureaucratic and organisational matters. Likewise, I want to thank Bernd Hofferberth and Klaus Hagdorn-Wittern for technical support.

Last but not least I want to express my sincerest gratitude to all those who were there for me in private life. Especially to my parents for everything and my husband Christoph for his support and all the great time. I am looking forward for more!